MODELLING HABITAT USE IN FOREST SONGBIRDS

A THESIS
SUBMITTED TO THE FACULTY OF THE GRADUATE SCHOOL
OF THE UNIVERSITY OF MINNESOTA
BY

Grant Canterbury

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
MASTER OF SCIENCE

March 1993

UNIVERSITY OF MINNESOTA

This is to certify that I have examined this bound copy of a master's thesis by

Grant Eric Canterbury

and have found that it is complete and satisfactory in all respects,
and that any and all revisions required by the final
examining committee have been made.

A. M. STARFIELD

Name of Faculty Adviser

Signature of Faculty Adviser

3-9-93

Date

GRADUATE SCHOOL

Table of Contents

I. Preface 1
II. Introduction. 3
1. Theoretical Background. 3
1.1 Does Population Density Indicate Habitat Quality? 3
1.2 Factors Affecting Bird Distribution. 6
1.2.1. Physical Features of Habitat. 6
1.2.2. Interspecific Interactions. 8
1.2.3. Non-environmental Factors. 11
1.2.4. Factors to Include in Models. 12
1.3. Presence/Absence vs. Density Measures of Abundance. 13
1.4. Data vs. Experience in Model-building. 14
1.5. Literature Review: Types of Models. 16
1.5.1. Ordination Methods. 16
1.5.2. Habitat Suitability Indices. 16
1.5.3. Multiple Regression Models. 17
1.5.4. Logistic Regression Models. 17
1.5.5. Rule-based Models. 18
1.5.6. Fuzzy Models. 19
2. Introduction: Birds and Vegetation of Itasca State Park. 21
2.1. Vegetation of the Park. 21
2.2. Bird Communities of the Park. 22
2.3. Bird Species Descriptions. 24
2.3.1. Ovenbird. 25
2.3.2. Red-eyed Vireo. 25
2.3.3. Pine Warbler. 26
2.3.4. Common Yellowthroat. 26
2.3.5. Black-throated Green Warbler. 26
2.3.6. Chestnut-sided Warbler. 27
2.3.7. American Redstart. 27
2.3.8. Least Flycatcher. 27
III. Methods & Analysis. 29
1. Description of Study Sites. 29
2. Vegetation Sampling Methods. 30
3. Breeding Bird Census Methods. 31
4. Data Analysis Methods. 33
4.1. Probability-of-Presence (POP) Models. 35
4.1.1. Building Rule-based and Fuzzy Models. 35
4.1.2. Building Logistic Regression Models. 36
4.1.3. Evaluating POP Models. 37
4.2. Population Density Models. 38
4.2.1. Building Population Density Models. 38
4.2.2. Evaluating Population Density Models. 38

IV. Results.
1. Ovenbird. 69
1.1. Probability-of-Presence Models. 69
1.2. Population Density Models. 70
2. Red-eyed Vireo. 71
2.1. Probability-of-Presence Models. 71
2.2. Population Density Models. 71
3. Pine Warbler. 72
3.1. Probability-of-Presence Models. 73
3.2. Population Density Models. 73
4. Common Yellowthroat. 74
4.1. Probability-of-Presence Models. 74
4.2. Population Density Models. 76
5. Black-throated Green Warbler. 76
5.1. Probability-of-Presence Models. 76
5.2. Population Density Models. 77
6. Chestnut-sided Warbler. 78
6.1. Probability-of-Presence Models. 78
6.2. Population Density Models. 79
7. American Redstart. 79
7.1. Probability-of-Presence Models. 80
7.2. Population Density Models. 80
8. Least Flycatcher 81
8.1. Probability-of-Presence Models. 81
8.2. Population Density Models. 81
V. Discussion 83
1. Do the Models Reflect Habitat Quality? 83
2. Statistical problems in Regression Analysis. 84
3. Choosing among Regression models. 85
4. Population Density vs. POP Models. 86
5. Roles of Data and Experience in POP Models 87
VI. Literature Cited. 90

Tables and Figures

Figure 1. Map of plot locations in Itasca State Park. 40
Figure 2. Summary maps of census plots. 41
Figure 3. Vegetation map of Sewage Lagoon Plot. 42
Figure 4. Vegetation map of Red Pine Plot. 43
Figure 5. Vegetation map of Mature Aspen Plot. 44
Figure 6. Vegetation map of Peace Pipe Plot. 45
Figure 7. Vegetation map of Checkpoint Plot. 46
Figure 8. Vegetation map of Mary Lake and Arco Lake Plots. 47
Figure 9. Examples of Habitat Data (Red-eyed Vireo) 48
Figure 10. Fuzzy Model Confidence Levels and Model Predictions (Red-eyed Vireo) 49
Figure 11. Sample Expert System Session. 50

Table 1. Habitat Types in Census Plots. 51
Table 2. Description and Ranges of Habitat Variables. 52
Table 3. Ovenbird Probability-of-Presence Models. 53
Table 4. Red-eyed Vireo Probability-of-Presence Models. 54
Table 5. Pine Warbler Probability-of-Presence Models. 55
Table 6. Common Yellowthroat Probability-of-Presence Models. 56
Table 7. Black-throated Green Warbler Probability-of-Presence Models. 57
Table 8. Chestnut-sided Warbler Probability-of-Presence Models. 58
Table 9. American Redstart Probability-of-Presence Models. 59
Table 10. Least Flycatcher Probability-of-Presence Models. 60
Table 11. Ovenbird Population Density Models. 61
Table 12. Red-eyed Vireo Population Density Models. 62
Table 13. Pine Warbler Population Density Models. 63
Table 14. Common Yellowthroat Population Density Models. 64
Table 15. Black-throated Green Warbler Population Density Models. 65
Table 16. Chestnut-sided Warbler Population Density Models. 66
Table 17. American Redstart Population Density Models. 67
Table 18. Least Flycatcher Population Density Models. 68

Preface

In this project, I build and evaluate models that use characteristics of the local vegetation to predict patterns of distribution and abundance for several songbird species that are common in the forests of northern Minnesota. To build the models I used data from several Breeding Bird Census plots in Itasca State Park, Clearwater County, where I measured vegetation and mapped bird territories during the summer of 1991. Field work was done with the aid of Dr. David Blockstein and the Field Ornithology class from the University of Minnesota Itasca Biological Station. The models are designed to apply to the mixed pine-deciduous forests of Itasca State Park and to similar forest habitats through northwestern and northcentral Minnesota.

Several themes are developed in this thesis, ranging from specific questions about the habitat preferences of individual species to more general issues concerning the advantages and disadvantages of different model-building methods. The major points are reviewed below.

- The models built here predict some index of species abundance (either the probability of the bird being present or local population density), and their accuracy can be evaluated by statistically comparing their predictions to independent data. However, the more fundamental purpose of these models is to give information about the importance of different habitats in maintaining the bird population, so that informed management decisions can be made. The usefulness of a model for land management is more closely related to its ability to predict habitat quality than to its ability to predict species abundance. These two concepts are related but not necessarily identical, and their relationship will be examined in more detail.

- Factors that can affect habitat use include the structural and floristic composition of the local vegetation, competitive and predatory interactions with other species, systems of territoriality, and large-scale population trends. The role played by each of these factors is examined. Each bird species is ecologically and behaviorally distinct and will be affected differently by these influences. Factors that are likely to

be important, to the birds in Itasca State Park in general and to the subject species in particular, are discussed below.

- The method used to census the bird population affects the way in which bird abundance is recorded, which in turn affects the type of model that can be drawn from the data. If the measure of bird abundance is qualitative (presence/absence at a site), the model suggested is one that predicts the probability of the bird being present or the percentage of a habitat type that is occupied by the species. Census methods that give a quantitative measure of abundance suggest a model that predicts population density. The advantages and disadvantages of the two modelling methods are evaluated in the text.

- The modeller has two sources available for information about the habitat use of a bird species: data collected in the field (habitat characteristics, bird distribution, etc.), and personal knowledge about the biology of the species (drawing on informal observations, studies in the literature, etc.). The first is quantified and can be used objectively to build the model, while the second is seldom sharply defined and requires personal judgement to incorporate it into the model-building process. Both are important, but they are used differently in different types of models. The interplay between data and experience is highlighted throughout this discussion.

Introduction

1. Theoretical Background

1.1 Does Population Density Indicate Habitat Quality?

The concept of "habitat quality" is present in the background of most studies that examine the abundance and distribution of animals, but it is not always discussed directly. Generally habitat quality is considered to be related to the carrying capacity of an area. However, the presence of a large population does not necessarily imply high survivorship or breeding success, which would indicate high-quality habitat. A population may be maintained in an inferior habitat by immigration without being able to reproduce successfully (Wiens 1981, Gibbs and Faaborg 1990). Van Horne (1983) defines habitat quality as a measure of mean individual "fitness" per unit area, in a formula incorporating population density, fecundity, and probability of survival for juveniles and adults. This measure is consistent with a view of high-quality habitats as areas that over the long term will satisfy a species' requirements for life, allowing successful reproduction as well as high population levels. Sites that have high habitat quality by this definition are also the sites that should have the highest priority for conservation in programs to maintain a species' population, so methods for determining habitat quality have pragmatic uses in conservation biology and wildlife management. Estimating habitat quality is then more important as a goal for a model than simply modelling the distribution of the population through space, which will not necessarily give important information for management.

An accurate assessment of habitat quality requires intensive study of reproductive success and survivorship in marked populations across a variety of habitats over several years (Van Horne 1983). Such comprehensive studies are rare, due to the difficulty of the undertaking. Somewhat less intensive approaches to evaluating habitat quality include the monitoring of physiological measures of well-being (body fat, etc.) and evaluation of food availability in an area. Generally, however, some index of population abundance has been used, with the assumption that areas with dense populations are also high-quality habitat in the sense of providing an opportunity for

successful breeding. Van Horne noted several factors that can cause the correlation between population density and habitat quality to break down. There may be variation in population density between years that is not related to the long-term quality of a particular habitat type (i.e. transient effects of predator density or climatic variations). Social dominance interactions between animals may also lead to substantial populations of subdominant individuals (juveniles or less dominant adults) accumulating in lower-quality habitats. For animals that can use a site year-round, winter mortality may be the main factor limiting population levels so that summer estimates of density have little relation to the year-round habitat quality of a site.

Thus, some caution is needed when using population density as a surrogate for information about habitat quality. With this caution in mind, it is interesting to examine whether the modelling of population distribution for migratory songbirds, as I have done in Itasca Park, can be used to evaluate the more complex concept of habitat quality. Several points could confuse evaluations of habitat quality for migratory birds. Annual variation in population density (e.g. Wiens 1981a) can complicate comparison of results from different years. Several British bird species (O'Connor, 1981) use a broad range of habitat types in years of high population density, but they are found only in specific habitats (presumably the preferred ones) when population levels are low. This is likely to be characteristic of American species as well, and it could prevent models from accurately identifying habitat preferences if they were built from frequency-of-occurrence data taken at high population levels. Non-territorial "floaters" may also accumulate in low-quality habitat if the high-quality areas are occupied by more dominant territorial individuals. This may also skew estimates of habitat quality if the floaters are counted in the population. Unmated territorial males, which may have been less dominant individuals forced into low-quality habitats, are likely to sing more persistently and be more detectable than mated individuals (Gibbs and Faaborg, 1990). Single-visit surveys that do not intensively study a population may thus underestimate population densities in favorable habitat.

It is possible to compensate for several of these confounding factors, however. If models are built from information about mapped territories, non-territorial birds are

unlikely to be included because they are not consistently present at a site and are less likely to be counted because they do not sing or show territorial defense behavior. The tendency to underestimate populations of mated territorial males, compared to actively singing unmated territorial males, is potentially more serious. This can be partially counteracted by making multiple visits to the site, so that the less detectable mated individuals are observed repeatedly and counted as present. Song playback can also draw out territorial behavior from inconspicuous individuals. If the breeding season is fairly synchronized (as it was in this study), any influx of subdominant juveniles into low-quality habitat does not begin to occur until late in the breeding season, and performing the census before the young fledge ensures that only adult birds are counted. Among migratory birds the breeding area is not used during the winter, so an early-summer census that covers the nesting period includes the season that is most critical in determining the quality of the breeding habitat (although the overall status of the population may well be controlled by factors on the wintering grounds).

There is also generally an inverse relationship between food abundance and the size of a territory. This has been demonstrated for numerous species (for a review see Smith & Shugart, 1987). Because food availability is an important factor in habitat quality, this suggests that, even in areas that are packed with territories, information on habitat quality might be gained by examining variations in territory size. This effect can partially offset the problem of habitat selectivity changing with population size (regardless of population levels, higher-quality areas should have smaller, more densely packed territories). However, if there is a minimum viable territory size (i.e. the "elastic disk" model of territoriality [Huxley, 1934]), a large population will eventually reach a point at which it is difficult for territories to be compressed further and territory size will not give information about habitat quality. Tompa (1962) studied an extremely dense island Song Sparrow population on an island, and showed that territorial males without mates had much smaller territories than mated males, probably indicating lesser competitive ability. Although there was no evidence for differences in habitat quality between territories in Tompa's study, his results do suggest that less dominant individuals could be forced into small territories in inferior habitat if space is at a

premium, thus reversing the usual relation between habitat quality and territory size noted above. Very high populations thus may prevent models based on density from accurately assessing habitat quality.

1.2. Factors Affecting Bird Distribution.

The factors that affect the distribution of birds range from physical features of the habitat and biological interactions with other species to influences that are not clearly part of the bird's environment (e.g. territorial behavior, regional population levels, etc.). Some factors are probably important influences on population distribution because they affect the quality of the habitat, while others do not seem to be related to habitat quality. Some of these factors can be measured easily and some cannot; this practical point is also important in deciding what information should be used in building a model.

1.2.1. Physical Features of Habitat. The physical aspects of the habitat, including the structure and species composition of the vegetation as well as landform features (cliffs, marshes, buildings, etc.) have a strong influence on the quality of a habitat for birds. The availability of important life requisites (e.g. nesting sites, foraging substrates, cover from predators) is controlled by these physical aspects of the environment, and the resulting variations in habitat quality are typically reflected in the distribution of the population through space. The measurement of physical habitat features such as these is usually straightforward, as they are typically long-lasting and easily accessible physical objects.

The local vegetation structure (density of shrubs, canopy closure, number of foliage layers, etc.) clearly influences the distribution patterns of birds. Abundant research has demonstrated this, going back at least to the work of MacArthur and MacArthur (1961), which correlated foliage height diversity with bird species diversity. James and Wamer (1982) have used more complex multivariate techniques to evaluate the relationship between species diversity and a larger array of structural habitat variables. These papers were concerned with relating habitat structure to the diversity of the bird community as a whole, rather than to the distribution of individual species.

Other authors have, however, examined the habitat use patterns of particular species across gradients of vegetation structure using local vegetation samples and multivariate ordination techniques (for example: James, 1971; Anderson and Shugart, 1974; Collins et al., 1982; Finch, 1989). All species have some variation in abundance that is affected by habitat structure, and different species will generally have differing criteria for which habitat types they will occupy.

The species composition of the vegetation has a more ambiguous effect on distribution patterns of birds. Some bird species definitely vary in their abundance between habitats dominated by different plant species (see, for example, Sherry and Holmes 1985). However, this factor is difficult to separate from vegetation structure because structural characteristics are often associated with particular taxonomic groups of plants (see Rotenberry 1985 for further discussion of this question). Conifers, for example, are clearly structurally distinct from deciduous trees, but there are also more subtle characteristics of structure (e.g., length of the petioles on the leaves, arrangement of leaves within the tree canopy) that vary between tree species and can affect a bird's foraging efficiency (Maurer and Whitmore, 1981; Robinson and Holmes, 1984). Insect abundance, discussed below, may also vary between tree species. Unfortunately, it is difficult to investigate the mechanisms by which these minor differences among plant species can affect bird distribution. It requires a sophisticated knowledge of a bird's foraging habits to determine whether or not the foliage structure of a tree is favorable to its feeding techniques, and measurement of insect populations is difficult. Trees that are clearly different in structure (conifers vs. deciduous trees, pine vs. spruce/fir) are most likely to support different bird communities. More similar tree species (red pine vs. white pine, sugar maple vs. white oak) are likely to have only subtle differences between their associated bird communities, and these differences will be difficult to detect.

The broad structure of the habitat can also have a significant effect on bird distribution. In continuous forests, the availability of widely dispersed habitat features such as exposed cliffs, bodies of water, standing dead snags, or large pines may not be represented accurately by the vegetation in a small area. A local vegetation sample

may not give useful information about habitat quality for birds with large territories such as hawks or woodpeckers, which may be keying in on the availability of a single localized nest site or the general vegetation type over a large area.

Another aspect of broad habitat structure is the effect of "edge" or the size of the forest patch. Two otherwise similar forest stands located in different surroundings (city park vs. continuous forest) would have distinctly different bird communities. Askins and Philbrick (1987) found that populations of many forest songbirds in a long-term census plot declined when the regional forest cover was reduced, although the plot itself had not been disturbed. Similar patterns have occurred in other long-term plots that have been cut off from surrounding forests by expanding suburbs in the past forty years (Askins et al., 1990). Many forest bird species in the eastern U.S. appear to be sensitive to the amount of edge habitat in forest fragments. Temple (1984) has found that populations of some forest species on small woodlots can be predicted better by using the "core area" of the forest (more than 100 m from the edge) than by using total forest area. In the case of larger bird species requiring forest habitat, an isolated patch of forest might not be large enough to support an individual. Edge effects can thus be important in affecting bird distribution.

1.2.2. Interspecific Interactions. Interactions with other species, including food availability, disease, predation, interspecific competition, and brood parasitism, can also affect population distributions. These factors will affect the habitat quality of a site to the extent that they are reliably associated with it, but the populations of other species are subject to fluctuations. Periods of intense predation pressure, food scarcity, or disease may be episodic, changing population levels temporarily without affecting long-term habitat quality. The measurement of these factors ranges from moderately difficult to impossible. Animals are generally mobile, inconspicuous, and less amenable to measurement than are plants and landforms, so population estimates are not precise. Even with an accurate census, the influence that an animal may have on the target species is usually complex and poorly understood. Models incorporating interactions with other species are probably not practically useful unless the abundance of the second species can be easily estimated and it has a clear and consistent effect on the

target species. If the effect of the second species is reliably dependent on some physical characteristics of the habitat, it may be just as well for practical purposes to use that habitat feature in the model instead of the species interaction. The use of the model in the field will be much easier if one is measuring, for example, the distance to the forest edge rather than the abundance of raccoons and cowbirds. Such a physical measurement may also give better information about long-term habitat quality than a census of predator or competitor populations, which might fluctuate dramatically from year to year.

Some of the negative effects of small stand size and "edge" habitat on forest birds are probably due to interspecific interactions. Birds characteristic of disturbed habitats (Starlings, House Sparrows, Robins, etc.) would also be more common in isolated patches and might competitively exclude forest-interior species. Factors such as nest predation by edge species such as raccoons and Blue Jays affect reproductive success more severely in small woodlots, especially for ground-nesting species (Wilcove, 1985). Brood parasitism by Brown-headed Cowbirds is also a more serious problem in isolated forest patches, where Cowbirds are more common. While many open-country species have developed adaptations to counter brood parasitism (building a new nest or walling off foreign eggs), forest-interior species are vulnerable when they come into contact with Cowbirds in fragmented forests (Brittingham and Temple, 1983). These interactions can significantly affect the distribution of area-sensitive species in fragmented habitats. Because these interactions are consistently associated with forest size, forest-edge areas can be fairly characterized as having poorer habitat quality for these species than otherwise similar continuous forests.

As noted above, food availability is an important determinant of habitat quality. Insect abundance may differ among tree species or structural habitats, and this could reasonably affect the habitat preference of insectivorous birds. Brush and Stiles (1984) found differences in seasonal abundance of leaf-dwelling insects between pines and oaks, and these differences were correlated with the breeding phenologies of the local songbirds. The size of Ovenbird territories is also related to the abundance of invertebrates in leaf litter (Smith & Shugart, 1987). However, measurement of insect

population levels is time-intensive, has several potential sources of error, (see Blenden et al., 1984), and may not indicate the actual availability of the resource to the birds. Changes in insect populations may occur from year to year and from week to week, so the relevance of isolated measurements to the long-term habitat quality of a site is questionable.

Interspecific competition is often difficult to distinguish from other factors in its effects on bird distribution. Two species with different habitat preferences may not overlap in their spatial distribution, but this does not necessarily imply that one species has any negative effect on the other. Differences in habitat use or foraging techniques that reduce the effects of competition may have originally evolved due to competitive displacement ("the ghost of competition past"), but this is difficult to prove (Connell, 1983). Wiens (1977) has noted that competition will be rare if populations are usually kept below carrying capacity by factors such as predation and climatic variability, and resources are not limiting. A controversial study of bird communities on Australasian islands (Diamond, 1975) does suggest that competition among similar species may be important in structuring communities on an evolutionary time scale, but the process is probably so subtle that it would be difficult to see in the course of a few years. Mountainspring and Scott (1984) attempted to detect competitive effects among Hawaiian forest birds, determining habitat preferences and factoring them out by examining partial correlations between species. Their extremely large set of census and vegetation data allowed them to detect several weak negative correlations, usually between exotic and native Hawaiian species. These apparent competition effects seemed to be sporadic and weak, although they could potentially lead to species replacement if integrated over many years. The relative rarity of negative correlations among native birds suggests that competition was not an active factor controlling their distribution prior to introduction of exotics. Perhaps overt effects of competition had been removed by competitive displacement over evolutionary time. One of the best demonstrations of competition among songbirds was given by Sherry (1979), who showed experimentally that the presence of Least Flycatchers depressed the abundance of American Redstarts. Flycatchers are particularly aggressive birds, and the similarity

between these two species in foraging style and habitat use probably made interspecific aggression more likely. In another example, Red-winged Blackbirds arrive early in the spring to occupy preferred territories in the center of marshes, but are later evicted to the margins by the larger Yellow-headed Blackbirds (Orians and Willson 1964). In general, the effect of interspecific competition on bird distribution is not clear, but it seems likely that it is most important in ecologically similar species that are not clearly separated by difference in habitat preference (and particularly among species that have only recently come into contact). Competition between evenly matched species is probably difficult to detect; in cases where competition has been convincingly documented one species has been clearly dominant. In most cases interspecific competition is appears to play only a minor role in determining bird distribution.

1.2.3. Non-environmental Factors. Several factors that govern the distribution pattern of a species are not clearly part of its external environment but are specific to the species' own biology and population structure. These include the species' system of territoriality and factors reflecting idiosyncrasies of its history, such as geographic range and overall population size. These factors are not related to the habitat quality of a particular site but can have an effect on population distribution - territoriality affecting local patterns, range controlling regional patterns, and population status affecting abundance and habitat selectivity.

Intraspecific territoriality is a common influence on the distribution of songbirds. Among migratory American species, the males typically arrive on the breeding ground in the spring, advertise territoriality by song, and defend their individual territories from neighboring males. Females then choose among the territorial males. The territories are used for courtship, nesting, and to supply food for the adults and young (Type A territoriality, as described by Hinde [1956]). Not all males defending territories are necessarily successful in attracting a mate. Although the mating status of territorial birds is useful for evaluating which habitat types are most favorable, information about it is also difficult to establish in the field without detailed observations and individually marked birds. Sometimes non-territorial "floater" individuals are observed as well. These are often young birds that could not establish

a territory and are not consistently present at a site during the breeding season (Stewart and Aldrich, 1951; Tompa, 1962; Carrick, 1963). Although territoriality is the common pattern in temperate songbirds, some species that feed on patchy, unpredictable resources or that have especially large home ranges do have less strict territorial behavior, and they may defend only the nest site.

Geographic location may also influence the local abundance of birds. The range of any species has been affected over time by its climatic tolerances, migration and colonization abilities, and past history, as well as by habitat type and interspecific interactions. At the most basic level, it is unlikely that a species will be present, even in ideal habitat, if the location is outside its usual geographic range. The location within the range may also affect abundance - sometimes populations are highest and most stable in the center of a species' range, where environmental conditions are likely to be most favorable. Marginal populations may be "sinks," rather than "sources," which are dependent on population overflow from the center for maintenance and are more variable from year to year (Wiens 1981).

Finally, the status of the regional or global populations of a species may affect its local population levels in ways that are not related to local conditions. A migratory population may change because of factors on the migration routes or wintering grounds (e.g. hunting, disease, habitat improvement or destruction), thus changing the number of individuals returning to breed (Terborgh, 1990; Askins et al., 1990). A decrease in population would probably lead to significant population decreases in marginal habitat but might not initially affect populations in the best habitats (Fretwell and Lucas, 1969; Kluyver and Tinbergen, 1953).

1.2.4. Factors to Include in Models. In general, interspecific interactions can be important determinants of habitat quality (the interaction with food organisms, in particular, is crucial). The variability of species populations combines with the difficulty of accurately measuring their population densities and their effects on the target species, so it is hard to directly incorporate interspecific interactions into a model of population distribution.

The non-environmental factors of territoriality, range, and population status are

more important for defining the parameters and range of applicability for a model than they are for predicting distribution or habitat suitability. The common system of non-overlapping breeding territories can be taken as a baseline for the construction of territory maps and models, and a species' range will indicate limits within which a model might be appropriately applied.

Physical features of habitat can be more effectively incorporated into models as predictive variables than can the other factors discussed above. They are easily measured, durable, and can affect the habitat quality of a site. Consequently, models of bird distribution have usually been "habitat models," using habitat characters to predict some index of suitability or abundance. The approach followed here is similar: habitat variables are used in the models, without explicitly including interactions with other animal species.

1.3. Presence/Absence Methods vs. Population Density Methods

The way in which the abundance of the birds is measured affects the type of model that can be built from the data. The simplest measure of abundance is a qualitative description of whether the species is present on the site being observed. This presence/absence value can be estimated with varying accuracy from several different census techniques. The simplest way of doing this is to make a single visit to a site and record whether the species is detected during a set time interval. There are several possible errors with this approach, including variations in weather, uncertainty about whether a detection of a distant bird should be scored as "present," and the possibility that birds could be present but remain undetected. A more accurate way of getting this information is by using the spot-mapping methods of the Breeding Bird Census (BBC) in conjunction with song playback to identify territories, making multiple visits to the site to construct territory maps, and scoring sites as occupied or unoccupied based on whether they fall within a territory boundary. Models built from presence/absence data will give as an output the probability of the species being present at the site (or, alternatively, the proportion of a habitat type that is occupied by the species).

The second method of measuring abundance is quantitative rather than qualitative, giving some scalar numerical value rather than a presence/absence score. Here again, estimates made with different methods can vary greatly in accuracy. Point counts such as those done in the Breeding Bird Survey (BBS) count the number of individual birds heard per stop, and the counts can be used as an index of the abundance of the species. This method is effective as a survey of population levels over wide areas, but it is not designed to examine local distributions and probably gives only a rough idea of population abundance near individual stops. The territory-mapping methods of the Breeding Bird Census (BBC), on the other hand, are poorer at detecting broad spatial trends in population but provide a more accurate measure of local population levels and the spatial distribution of territories over short distances. The number of territories within a given area can be used as a good measurement of local population levels. BBC results have the advantage of giving actual estimates of population density rather than vague indices of abundance. Models built from quantitative abundance data such as these will give as an output the predicted abundance level at a site.

The choice of which method to use involves a tradeoff between the labor required in the field and the relevance of the data to determination of habitat quality. In this project I use data from territory maps to give values for both presence/absence and local population density. I expect the presence/absence data to be more accurate (see Methods & Analysis). In general, presence/absence data should require less labor to obtain than population density data. However, the previous discussion about habitat quality (§2.1) indicates that territory size and population density are expected to correlate with the suitability of the habitat. Presence/absence determinations, on the other hand, may have little relationship with habitat quality when population levels are high and most of the available habitat is being used. The use of both modelling methods in this analysis allows a comparison of their effectiveness in discriminating between used and unused habitat.

1.4. Data vs. Experience in Model-building.

The information used in building a model comes both from data collected in the

field and from personal knowledge about the biology of the species. Field data, including the values of habitat variables and information about the distribution of birds, are selected pieces of information abstracted from the environment of the study site. Data are explicitly recorded and can be used in a straightforward way to build statistical models. In model-building methods where personal judgement plays an important role, the presence of field data is valuable because it gives an objective "reality check" against possible mistaken ideas. It is also easier for qualitative ideas such as "Ovenbirds are more common in closed forests" to be expressed in quantitative, testable form when a graph shows the relation between abundance and canopy cover. Independent verification data can provide a clear criterion against which the predictions of models can be judged, and so they are useful in evaluating model accuracy.

Information coming from experience is less easily defined. This is usually not set down in explicit form, and expressing it completely may be impossible. The idea of optimal habitat derived from experience is likely to be a gestalt impression of the appearance of areas with good habitat, rather than a quantitative description. Habitat descriptions in the literature, past observations of the species, communications from other observers, etc., all contribute to this information. This information is important in the initial selection of species to be studied and in setting the parameters of the investigation. Judgement is obviously needed for incorporating experience into a model, either for selecting the variables to be used in a regression or for specifying a response curve for the relation between habitat type and bird abundance. The incorporation of personal judgement into a model cannot be done mechanically. Neither can it be completely avoided - if nothing else, the habitat variables to be measured must be chosen from an among an infinite number of possible variables that could be used. Consequently, all models are partly subjective. This subjectivity means that the work of different analysts is likely to lead to different models. However, it also can allow potentially important knowledge to be used from beyond the limits of the dataset, increasing the model's predictive power, and so it is not necessarily harmful.

1.5. Literature Review: Model Techniques.

The various types of models that have been used to evaluate bird habitat all emphasize the measurement of physical features of the environment. However, the approaches use different measures of population abundance, and they incorporate data and experience into the model-building process in different ways. The major methods for examining bird-habitat relationships are summarized below.

1.5.1. Ordination. Many studies have used principal components analysis (PCA) or similar ordination methods to examine habitat use (e.g. James, 1971; Finch, 1989; Sherry and Holmes, 1985; Collins et al., 1982). Ordination techniques are used to compress the multi-dimensional variation in habitat characteristics into a smaller number of dimensions that express the variation in habitat as efficiently as possible. The sites that are used by the species can then be plotted on the ordination axis or axes, allowing differences in habitat use between species to be visualized easily. This method is often used with ecological studies of niche separation between species, community structure, or habitat preferences. It is essentially a simplified visual representation of the data. The main use of ordination is as a description of habitat use patterns - it is difficult to get predictions about habitat use from these methods.

1.5.2. Habitat Suitability Indices. The Habitat Suitability Index (HSI) is an official U.S. Fish & Wildlife Service method of habitat description. The method sets up univariate plots of habitat suitability against various habitat characteristics, with suitability values ranging from 0.0 (unsuitable habitat) to 1.0 (optimal). The suitability values for the several variables are then combined in a modified geometric mean or other arithmetical combination to give an overall habitat suitability rating. There is apparently no provision for incorporating interactions between variables. HSI models have been recently criticized for a model-building process based only on expert opinion and literature review without direct involvement of field work (Lancia et al., 1982). This method thus relies almost exclusively on experience, without using field data in the model-building process. Often evaluation of the models has been based only on expert opinion of their accuracy, which justifies them by a circular argument because the same experts were typically consulted when building the model in the first place.

The model does give an output that can be field-tested, although the suitability value is a general measure of habitat quality rather than a direct prediction of probability of presence or population density. Validation of HSI models with independent data has been done a few times, typically by censusing a number of sites and comparing their ranking by actual abundances with their ranking by predicted HSI values (Bart et al., 1984; Cole and Smith, 1983).

1.5.3. Multiple Regression Models. If some quantitative measure of population density is available for each habitat point, multiple regression methods can be used to develop a predictive model that will estimate population density given values for the habitat variables (e.g. Laymon and Barrett, 1986). One to several habitat characteristics are chosen to be used as predictor variables. (If non-linear effects are likely, the square of the variable can be entered into the model to give a curvilinear model.) The model produced by multiple regression is an equation that uses coefficients assigned to each predictor variable plus the intercept, to give a predicted value for population density. The relation between the predicted population density and the predictor variables is a best-fit line or response surface that minimizes the variance between actual observations and predicted values. Knowledge about which variables are likely to be biologically important should be used when choosing variables for a model. Automatic stepwise regression, giving the "optimum" choice of variables, is a risky process and frequently fails in selecting the most influential variables (James & McCulloch, 1984). The choice of variables is the major way for personal judgement to be used in building multiple regression models - once that decision has been made, the regression procedure automatically determines the response surface from the data.

1.5.4. Logistic Regression Models. Multiple logistic regression is similar in many ways to multiple regression but works with presence/absence data to produce a probabilistic model rather than a quantitative one (e.g. Capen et al., 1986; Brennan et al., 1986). Datasets analyzed with logistic regression have several independent predictor variables (habitat characteristics) and a single categorical dependent variable (presence or absence of the species). One to several variables are chosen for the

model, which may include quadratic terms. Automatic stepwise methods for selecting variables should be avoided, as in multiple regression. The model produced by the process is an equation into which one can substitute the values of habitat variables for a location to produce the predicted probability that the species is present on the site.

1.5.5. Rule-based Models. Regression analysis does not necessarily give an accurate reflection of the way in which the bird population evaluates habitat. There is no a priori reason to believe that the response of a bird population to a habitat variable should be linear or quadratic in nature. Perhaps birds use relatively simple rules of thumb in evaluating habitat (e.g., dividing the woods into "open forest" and "closed forest," to use the rule-based Red-eyed Vireo model from Table 4 for purposes of demonstration). A rule-based model mimics this way of looking at habitat. The environment is classified into several different habitat types, using simple rules to define sharp cutoff points between types (e.g., "If Canopy Cover < 40% the habitat is Open Forest"), and the probability of the species being present is assigned for each habitat type (e.g. "In Open Forest there is a 15% probability of Red-eyed Vireos being present; otherwise there is a 70% probability"). This method can be partly or completely independent of field data — it is possible to assign cutoff points and probabilities of presence on the basis of personal opinions without formally collecting any data at all. Data are extremely useful as a guide, showing which cutoff points and probabilities of presence are plausible and which ones are not. Ultimately, however, the selection of the model is a matter of personal judgement. If one feels that the data are misleading or don't give enough information, one is free to ignore them and develop different rules. Starfield et al. (1989) and Starfield (1990) discuss related applications of rule-based models.

The rule-based modelling method presented here gives as an output a probability that the species is present. Its results are thus comparable with those of logistic regression models. An advantage of the rule-based method is that it can easily be incorporated into an expert system. Expert systems are interactive computer programs that utilize facts supplied by the user, in conjunction with specified decision rules, to solve a problem within the program's defined area of competence (see Coulson et al.,

1987; Giarratano and Riley, 1989). A rule-based model, built as described above, can be translated directly into decision rules for an expert system, which can then be used to predict probabilities of presence for the bird species. Other modelling techniques, which predict continuous variation in abundance rather than sharp discontinuities, are not easily adapted to the "If-Then" logic generally used in the decision rules for expert systems. The major advantage of an expert system is its ability to describe to the lay user the rules that it is using to make a decision and the biological rationale used to arrive at those rules. Figure 11 shows a sample expert system session, for which a rule-based model developed in this project was translated into expert system format and supplemented with explanatory comments. This kind of output is more easily understandable than a regression program, which gives a predicted value and a few explanatory variables that may have no clear explanation for their presence other than good significance values. The question of whether one approach is practically superior to the other will be addressed later in the paper.

The rule-based/expert-system approach is virtually untried in the habitat modelling literature. Marcot (1986) does describe a rule-based expert system for predicting habitat use by birds, but this is done primarily as a demonstration of their feasibility, in the course of a theoretical discussion of expert systems. No other applications of this approach to habitat modelling are present in the literature.

1.5.6. Fuzzy Models. The field of fuzzy logic has developed as a set of mathematical methods for quantifying and reasoning with imprecise or unreliable information. Fuzzy logic relaxes the classical assumption that statements must be completely true or false, allowing statements for which there are varying degrees of confidence in their truth (Zadeh, 1965; Schneider and Kandel, 1991). Some of the simpler applications of this body of theory can be used to modify the rule-based modelling method described above.

It is probably not realistic to assume that the birds' perception of habitat changes abruptly from one type to another when some habitat variable moves across a threshold. It is more likely that there is a gradual transition between types, during which there is some uncertainty about which type is present. To use the Red-eyed Vireo example

given in the discussion of rule-based models above (§1.5.5) , the transition point between Open Forest and Closed Forest may not occur abruptly at 40% Canopy Cover. For the fuzzy model represented in Figure 10, it is assumed that the confidence level in the statement "This is Open Forest" is 1.0 (True) at 30% Canopy Cover or less, and 0.0 (False) at 50% Canopy Cover or more. Between 30% and 50% Canopy Cover, the confidence level has intermediate values. For simplicity it's assumed that the relation between confidence level and Canopy Cover is linear between 30% and 50% Canopy Cover, although there are many potential shapes for this curve. The predicted probability of the species being present is the predicted probability for each habitat type multiplied by the confidence level that the site actually is in that habitat type, summed over all habitat types. If the probability-of-presence values that were specified in the rule-based model are left the same (If Open Forest: 15%; If Closed Forest: 70%), then our predicted probability of presence will be 20% when Canopy Cover is less than 30%, will rise linearly to 70% as Canopy Cover increases from 30 to 50%, and will equal 70% when Canopy Cover is above 50%.

This method allows judgement to be used extensively in model-building, while also allowing continuous variation in the predicted abundance value. Fuzzy models such as this can be incorporated into some expert system programs, although many others require the sharp category distinctions of ordinary rule-based modelling. While fuzzy logic has been developed theoretically and has had some practical applications in industry (see Schneider and Kandel, 1991), these ideas have so far not penetrated to the literature on habitat modelling.

2. Birds and Vegetation of Itasca State Park.

At this point, an introduction to the specific situation being modelled will be useful. Below I describe the vegetation types and bird communities characteristic of Itasca Park and review the biological literature on the bird species studied for this thesis.

2.1. Vegetation of the Park. All field work for this project was done in Itasca State Park, which is located in Clearwater and Hubbard Counties in north-central Minnesota. This area straddles the ecological boundaries between the northern aspen-conifer forests, the southern maple-basswood forests, and the western tallgrass prairie. Itasca State Park is a large tract of mixed coniferous-deciduous forest surrounding Lake Itasca, which is the headwaters of the Mississippi River. The forest is nearly continuous, broken only by small wetlands, a few roads and hiking trails, and local areas developed for recreational use. There has been no extensive logging since the early part of this century. The upland forest habitats within the park range from stands of mature red pine (Pinus resinosa) and white pine (P. strobus) to deciduous forests dominated by quaking aspen (Populus tremuloides), paper birch (Betula papyrifera), sugar maple (Acer saccharum), red oak (Quercus rubra), and bur oak (Q. macrocarpa). Bog forests in lowland areas of the park are dominated by black spruce (Picea mariana) and tamarack (Larix laricina). Minor constituents of the park forests include balsam fir (Abies balsamea), white spruce (Picea glauca), red maple (Acer rubrum), basswood (Tilia americana), ironwood (Ostrya virginiana), and black ash (Fraxinus nigra). Much of the forest in the park also has a dense shrub layer of beaked hazel (Corylus cornuta). The mixed deciduous forests usually have closed canopies 10-20 m in height, but mature aspen trees typically reach 20 m, and mature pines average 20-25 m tall. The canopies of aspen and pine are more open than those of the other trees and these species are often associated with dense layers of smaller trees or shrubs because more light is available at the lower levels of the forest.

Most of the mature stands of pine and aspen originated during the nineteenth century, in the aftermath of fires that burned large areas of the park. Before modern

fire control efforts, the park forests burned frequently at an average return interval of 22 years (Collins et al., 1982). However, fires have been suppressed in the park since the 1920's, so the understories in forest habitats are now typically dominated by tree species that are shade-tolerant and sensitive to fire, such as sugar maple, ironwood, and basswood. Pine regeneration is almost nonexistent, both because of the current lack of forest fires that could open up the canopy, and because pine seedlings are heavily browsed by deer, which have greatly increased their population in the past century. If current conditions hold, it is likely that the forest will become dominated by a maple-basswood vegetation over the next century, as old pine and aspen trees die off without reproducing.

2.2. Bird Communities of the Park. Here it will be useful to summarize the important background knowledge about the bird communities of the park, including general ornithological information as well as some observations from my field work. Approximately 45 bird species occur regularly in the upland forests of the park during the summer breeding season. The majority of these are migratory summer residents that spend the winter in the Neotropics or the southern United States. Permanent residents include grouse, jays, finches, chickadees, nuthatches, and some woodpeckers. The birds are all typical forest or forest-edge species; although there is prairie within 50 miles, no characteristic grassland birds are common within the park. Many of the forest-specialist species in the park are near the western boundaries of their range. Some changes in the bird communities of the park over the past thirty years have been attributed to variability in populations near the edge of their geographical ranges (Parmelee, 1977).

The migratory species return during May as the trees leaf out and insects become abundant. Male birds begin to defend territories immediately upon returning, and courtship and nesting take place during May and June. Fledglings begin to leave the nest in mid-July. Territorial defense diminishes rapidly through July and has virtually ceased by August. Fledglings and adults are nomadic during the late summer, often travelling in family groups or mixed-species flocks without territorial activity.

Year-round residents follow a similar pattern but can start breeding a month or more earlier, so the young have often already fledged and territories have broken down by June.

Variations in vegetation structure clearly affect the distribution of many species in the park, as is discussed in more detail below. The difference between pines and deciduous trees is significant to some species as well. Broad-scale habitat structure also has a few effects on bird distribution. Large pines are often widely separated in the forest, and they were not always represented in the localized vegetation samples that I used (see Methods & Analysis), but they are important habitat features for Pine Warblers and perhaps for Black-throated Green Warblers. A small marsh may also allow a Common Yellowthroat territory to extend a short distance into an otherwise unsuitable forest habitat nearby.

The continuous, unfragmented forest of the park allows high populations of those bird species that require forest-interior habitats. The species that are characteristic of fragmented forest and areas near human habitation are much less common within the park than they are in nearby farmlands and woodlots. In particular, the parasitic Brown-headed Cowbird is uncommon within the park. While adult cowbirds were present at low densities throughout the park, I observed only two cowbird fledglings during two summers of field work and no fledglings at all were seen in the mature forests that dominate the park.

Intraspecific territoriality is common among birds in the park. Most species show Type A territoriality, as described above (§1.2.3), with males defending non-overlapping territories that supply all of the resources needed for nesting. This mode of territorial behavior is taken as a baseline for the distribution maps developed from the census data. A few species that feed on patchy, unpredictable resources (swallows and finches) do deviate from the typical Type A pattern, and defend only the nest site. Species with particularly large home ranges (hawks and woodpeckers) may have partially overlapping territories because of the difficulty of defending the borders. Territory sizes vary greatly among breeding birds in the park. Broad-winged Hawks and Pileated Woodpeckers probably have territories covering tens or hundreds of

hectares. Smaller woodpeckers defend 2-5 ha. Passerine species usually have smaller territories: the thrushes and larger flycatchers defend 1-2 ha, while the commonest songbirds (warblers and vireos) average 0.3-0.4 ha. The Least Flycatcher has the smallest territories of all, typically 0.2 ha in size.

Interspecific territoriality is not obvious from the arrangement of territories that I observed. Cases where territories of two species do not overlap are explainable by differences in habitat preferences (i.e., Common Yellowthroats prefer open marshes while Ovenbirds prefer dense forests, so the lack of overlap in their territories is unlikely to be explained by competition). Interspecific competition might be expected among species pairs that are similar in foraging strategy and habitat use. Species pairs in the park that show such similarities include Ovenbird/Veery; Pine Warbler/Black-throated Green Warbler; American Redstart/Least Flycatcher; American Redstart/Chestnut-sided Warbler. The territories of all of these species pairs overlap extensively, and there is no apparent interspecific territoriality. The birds appear to select territories without considering the presence of other species.

The eight bird species chosen for detailed study were all typical migratory passerines with well-defined territories. The species included: Ovenbird, Red-eyed Vireo, Pine Warbler, Common Yellowthroat, Black-throated Green Warbler, Chestnut-sided Warbler, American Redstart, and Least Flycatcher.

2.3. Bird Species Descriptions. I will briefly summarize the geographical range, physical and ecological characteristics, and habitat preferences of the eight bird species studied, as described in the literature. Excellent illustrations and range maps are available in Peterson (1980). General information on foraging methods, breeding biology, and habitat preferences are taken from Ehrlich et al. (1988). The habitat use patterns of warblers in Itasca State Park have been specifically investigated with multivariate ordination techniques and cluster analysis by Collins et al. (1982) and Collins (1983a). Habitat use distributions of several of the species studied here were also analyzed by Sherry and Holmes (1985) in their study at Hubbard Brook, New Hampshire.

2.3.1. Ovenbird. This species is a large, ground-foraging warbler, similar in general appearance and feeding methods to a thrush. Ovenbirds are abundant in mature forests throughout southern Canada and the eastern United States, and they winter through Central America and northern South America. Ovenbirds feed by gleaning insects from the forest floor. These are ground-nesting birds, building a cryptic roofed nest out of dead leaves and forest litter. They are vulnerable to cowbird parasitism, and ground-nesting species such as this are at particular risk for nest predation (Wilcove 1985), although Ovenbirds' excellent nest camouflage undoubtedly helps to protect them. The birds occur in deciduous forests and sometimes in pine forests. Smith and Shugart (1987) found that Ovenbird territories were smaller in areas where prey density in the litter layer was high, and that a high density of litter invertebrates was correlated with habitat characteristics such as canopy closure, openness of ground cover and deciduous-dominated vegetation. Noon et al. (1980) surveyed Ovenbird habitat use across a range of geographical areas and found that they consistently prefer forests of tall trees with closed canopies. In the southern and central parts of their range Ovenbirds seemed to prefer deciduous forests, but they were more associated with conifers on sites in Michigan. Collins et al. (1982), in their study of warblers in Itasca, found that Ovenbirds were most common in mature forest but had no strong preference between deciduous and coniferous areas. Sherry and Holmes (1985) found Ovenbirds to have a slight, non-significant preference for deciduous vegetation on their New Hampshire sites. Ovenbirds are also strongly affected by habitat fragmentation and have poor breeding success near forest edges (Terborgh 1990; Gibbs and Faaborg 1990; Temple 1984).

2.3.2. Red-eyed Vireo. This is a species of the forest canopy that feeds on insects and caterpillars by gleaning them from plant surfaces or hovering to pick them from less accessible leaves. It is abundant throughout the eastern deciduous forest and is less common in the conifer forests of southern Canada and the northwestern U.S. The birds migrate to the western Amazon Basin for the winter. Red-eyed Vireos build a hanging nest in a shrub or small tree and are vulnerable to brood parasitism by cowbirds. They are not strictly forest-interior birds, being found in open woodland and

wooded suburbs as well as closed forests, but they appear to have suffered population declines in forest fragments isolated by urban development (Askins et al. 1990). Sherry and Holmes (1985) found that Red-eyed Vireos preferred deciduous-dominated sites on their New Hampshire census plots.

2.3.3. Pine Warbler. These are small, plain warblers that feed and nest almost exclusively in the upper canopies of large pine trees. They breed in pine forests throughout the eastern U.S. and winter from the southeastern U.S. to the Caribbean and Mexico. They feed primarily on insects, taking them from the bark of pine branches or clusters of needles. Their nests are only occasionally parasitized by cowbirds. Their strong preference for pine-dominated habitats is confirmed by Collins et al. (1982), who found Pine Warblers to be more strongly associated with large conifers than any other warbler in Itasca Park. Schroeder (1982) suggests that the presence of red pine is the most important factor for this species and that white pine is unsuitable habitat.

2.3.4. Common Yellowthroat. These warblers are wren-like in shape, with relatively short wings and large feet that adapt them to moving through dense ground cover. They are ground-dwelling birds, occurring in marshy or boggy areas, overgrown fields, or woodland edges. Yellowthroats feed on insects, picking them from the foliage of herbs or shrubs. Nests are typically built in low shrubs, and are frequently parasitized by cowbirds. They breed throughout the continental U.S. and southern Canada and winter from the southern U.S. to Panama. Collins et al. (1982) found that Yellowthroats in Itasca Park are strongly associated with dense ground cover and an open canopy.

2.3.5. Black-throated Green Warbler. These are canopy-dwelling birds that typically occur in coniferous or mixed forests, or occasionally in deciduous areas. They feed on insects gleaned from tree foliage. These warblers breed through the northeastern and northcentral U.S. and southern Canada, wintering in Central America. Their nests are usually built high in a conifer and are occasionally parasitized by cowbirds. According to Collins et al. (1982), Black-throated Green Warblers in Itasca Park are intermediate in habitat-use patterns between Pine Warblers and Ovenbirds,

with a concentration in closed forests and a moderate preference for conifer-dominated sites. Sherry and Holmes (1985) also found a slight, statistically non-significant tendency for Black-throated Green Warblers to occupy coniferous sites. Collins (1983b) surveyed habitat use for this species over a wide geographic area and found that the species used several different vegetation types, ranging from mature pine to spruce-cedar to maple-birch forests.

2.3.6. Chestnut-sided Warbler. These birds are characteristic of thickets and open, successional woodland. Their population has increased greatly over the past century as mature forest has been converted to brushy habitat. The warblers breed through the northeastern U.S. and southeastern Canada and winter in Central America. Like most of the other warblers, they search through tree and shrub foliage to feed on insects. The nests, which are typically built in a low shrub, are frequently parasitized by cowbirds, but the warblers have evolved some defensive egg-rejection behaviors. Collins et al. (1982) showed that Chestnut-sided Warblers occur in moderately open forest habitats (though not as open as for Yellowthroats), with a slight preference for small trees and deciduous-dominated areas.

2.3.7. American Redstart. These brightly colored warblers are characteristic of successional deciduous or mixed forest, though they occur in more mature stands than do Chestnut-sided Warblers. They feed actively on flying and foliage-dwelling insects, hovering and fanning their tails to dislodge prey items from leaves. Redstarts have broad, flycatcher-like bills with pronounced rictal bristles that aid in capturing flying insects. They range through the eastern and northwestern U.S. and southern Canada, wintering from Mexico to northern South America. Nests are built in small deciduous trees and are frequently parasitized by cowbirds. Collins et al. (1982) found Redstarts primarily in deciduous forests, on sites that were intermediate in canopy cover between those occupied by the Chestnut-sided Warbler and deep forest species like the Ovenbird. Sherry and Holmes (1985) also found that Redstarts favor deciduous vegetation over conifers and noted that they were less likely to establish territories in areas heavily used by the competing Least Flycatchers.

2.3.8. Least Flycatcher. Least Flycatchers typically occur in open deciduous

or mixed woodlands. They feed by sallying out from a perch to capture small flying insects or by hovering to glean insects from leaves. These bird usually feed and nest in the lower subcanopy of the forest. They breed through the northern U.S. and southern Canada, and winter through Mexico and Central America. Least Flycatchers have a system of territoriality that is unusual among Minnesota birds. Males defend mutually exclusive territories as the warblers and vireos do, but territories tend to be clumped together rather than being evenly spread through the appropriate habitat. This leads to the formation of "colonies" of 5-15 pairs of birds. Sherry and Holmes (1985) identify this as a noyau form of social organization. Because flycatchers are often aggressive toward members of other species as well as their own, individuals in a noyau colony may be better protected against predation and interspecific competition than are solitary individuals. Sherry (1979) has described Least Flycatchers aggressively attacking the ecologically similar American Redstarts and preventing Redstarts from establishing territories within flycatcher colonies. Sherry and Holmes also identify a strong preference by Least Flycatchers for deciduous areas on their study plot. Breckinridge (1956) examined the habitat characteristics of sites used by Least Flycatchers in an upland oak forest in southern Minnesota and concluded that an open subcanopy (8-30 feet in height) was a critical factor required by this species. This is consistent with the general perception of this species as an open-woodland bird, but the colonial habit of the species can make it difficult to detect habitat preferences. The abrupt change in abundance between sites within and outside a colony appears to suggest corresponding differences in habitat quality between the sites, but this implication is probably misleading. Sherry and Holmes note that some of their flycatcher colonies drifted in location over several years, which indicates that the unused sites were also probably suitable habitat.

II. Materials, Methods, & Analysis

1. Description of Study Sites

For this research project seven Breeding Bird Census plots were set up within the park, covering a total of 52 ha (Figure 1). The forests on the census plots range from mature pine to aspen-birch to maple-oak forests, representing the major upland forest types found in the park. In addition, several unforested marsh/bog habitats are present on the plots. Each plot is described briefly below, and the habitat characteristics of each plot are roughly summarized in Table 1.

Each census plot was surveyed using tape measure and compass. A grid of 50 x 50 m squares oriented to magnetic north was measured out and each grid intersection was marked with colored flagging tape. North-south grid-lines were marked with letters and east-west lines were marked with numbers so that the position of any point in the plot could be located in the field by reference to nearby grid markers.

The Sewage Lagoon plot (Figure 3), covering 10.25 ha, is located on the hillside north of the park sewage lagoon (47°13′57″ N, 95°10′33″ W). This plot contains two distinct vegetation types. The northern half of the plot includes mature stands of aspen and white oak, with a supercanopy of large white pines. The southern half was cleared of young trees and burned in 1972 as part of an unsuccessful pine regeneration experiment, and is the only area in the census plots that has been subject to major human disturbance in this century. The current vegetation includes an extremely dense growth of young aspen (5-10 m tall) and hazel, as well as scattered large white pines that were left standing when the forest was cut.

The Red Pine plot (Figure 4), covering 10 ha, is located on the Wilderness Drive of the park (47°11′52″ N, 95°14′37″ W). The major vegetation type is mature red pine forest with scattered white pines in a 20-25 m canopy, with an understory of maple. It also includes a 1.5-ha stand of deciduous forest dominated by aspen, a few canopy gaps with hazel thickets, and a 1-ha sedge marsh.

The Mature Aspen plot (Figure 5), covering 10 ha, is located near the intersection of the Ozawindib and Red Pine Trails, south of Arco Lake in the southern

part of the park (47°09′51″ N, 95°10′10″ W). The vegetation is mostly mixed deciduous forest, including mature aspen and birch, young sugar maple, and red oak. There is a stand of medium-sized red pines (c.15 m tall) in the southern part of the plot, and there are large pines scattered through the deciduous areas. The plot contains a 0.5-ha pond covered with reeds and a 0.5-ha sphagnum bog.

The Peace Pipe plot (Figure 6), covering 10 ha, is located southeast and across the road from the Peace Pipe overlook above the east arm of Lake Itasca (47°12′25″ N, 95°10′00″ W). The major vegetation type is mixed deciduous forest, including some mature aspen, maple, and oak. There are also stands of red pine and white pine and several canopy openings with dense hazel thickets. Several open marshes border the plot, and there are a few swampy forested hollows within the plot.

The Checkpoint plot (Figure 7), covering 6 ha, is located southeast of the south entrance to the Pine Ridge campground (47°13′25″ N, 95°10′20″ W). There is a stand of red pine at the west end of the plot, and a closed-canopy maple forest at the east end. Most of the plot is fairly open mixed deciduous forest, with scattered large pines and a dense shrub layer of hazel.

The Mary Lake plot (Figure 8), covering 2.75 ha, is located across the road from the southern tip of Mary Lake (47°11′00″ N, 95°10′45″ W). It is almost completely covered by mature red pine with a maple understory. There is a more open area dominated by deciduous vegetation at the eastern end.

The Arco Lake plot (Figure 8), covering 3 ha, is located east and across the road from Arco Lake (47°10′10″ N, 95°10′35″ W). Most of the plot is mature aspen with a maple understory and a thick layer of hazel. Only a few large pines are present, all at the east or west ends of the plot. There are shrubby clearings on the hillside at the east end and a small pond in the center of the plot.

2. Vegetation Sampling Methods

The vegetation in the census plots was measured from July 15-30, 1991. The structure of the vegetation during this period was not obviously different from what it had been in June, the most important part of the breeding season. Vegetation sampling

points were located at alternate grid intersections in a regular diagonal pattern, so that each sampling point was c. 71 m from its nearest neighbor (Figure 2). I sampled only half of the grid intersections to minimize the chance of two sampling points being included in the same bird territory. A total of 154 vegetation sampling points was measured over all census plots.

I used a modified version of the James & Shugart vegetation measurement technique (James & Shugart, 1970; James, 1978), which describes the vegetation within a circular plot of 11.2 m radius (0.1 acre = 0.04 ha). Trees were counted by species and size class, and shrub density was measured following the standard protocol above. The canopy cover measurements were expanded to allow a number of separate cover percentages to be calculated (ground cover; total canopy cover; subcanopy cover (< 10 m high); shrub-layer cover (< 5 m high); total canopy cover and subcanopy cover for each tree species). In addition, the approximate distances from the center of the circle to the nearest open wetland and to the nearest large pine were estimated, as these are important habitat characteristics that might not be represented within a plot of this size.

Table 2 lists the major habitat variables that were measured at each point. The observed ranges of each variable are also incorporated in this table, including the lowest value, highest value, and the 5th and 95th percentiles. This information is given to specify the range of applicability of the models derived from the data. Locations that have a vegetation structure or tree species composition significantly different from the training data are likely to have bird communities different from those predicted by the models given here. The models should be used with caution on sites that are outside the range from the 5th to 95th percentiles for these variables.

3. Breeding Bird Census Methods

Bird territories were mapped during June and early July of 1991. I was assisted in conducting censuses by Dr. David Blockstein and the students in the Field Ornithology class at the University of Minnesota Biological Station. Teams of 2-4 students worked on the Mature Aspen, Sewage Lagoon, and Red Pine plots. Many of the students were initially unfamiliar with bird identification, but Dr. Blockstein and I

assisted them on these plots while they were learning identification skills and the spot-mapping procedure. Because I worked extensively on all seven plots, differences between plots due to inter-observer variability were minimized. The remaining four plots I surveyed by myself, usually without other assistance.

The standard spot-mapping censusing technique was used (Robbins, 1978; Cornell Laboratory of Ornithology, 1989). We made eight visits to each plot, during the early morning (5-8 AM) or early evening (7-9 PM) when birds were most vocal. The locations of bird sightings, singing males, simultaneous-singing observations, and nest-defense behavior were recorded on the plot map for the appropriate species. A supplementary method was also used to determine territorial boundaries; I played recorded songs of the most common species (Red-eyed Vireo, Ovenbird, Black-throated Green Warbler, Pine Warbler, Chestnut-sided Warbler) at each vegetation sampling point in all plots to elicit territorial behavior from resident males and clarify whether the point was contained within a territory.

At the end of the season the species maps were examined to locate territory boundaries. The determination of territory boundaries was made by me, in consultation with any other individuals who had made observations on the plot. My involvement in censusing and determining territory boundaries on all plots minimized the effects of inter-observer and inter-analyst variability (Verner and Milne, 1990). Assignment of territory boundaries was based on repeated sightings in the same location, simultaneous singing between conspecific males, and other territorial behavior recorded in the field. Potential territories based on less than three observations were considered to be unreliable and were excluded to avoid recording chance detections of transient, nonterritorial individuals. A final map of each plot was drawn up for each species, showing the territorial boundaries of each male. The species chosen for further analysis were those that had fairly small territory sizes (so that one territory would not be likely to overlap two vegetation sampling points, which would complicate data analysis by reducing the independence of data points) and were common enough to give a fair idea of their habitat preferences. The best species for analysis by these criteria were Least Flycatcher, Red-eyed Vireo, Ovenbird, Pine Warbler, Chestnut-sided

Warbler, American Redstart, Black-throated Green Warbler, and Common Yellowthroat.

4. Data Analysis Methods

Before proceeding with data analysis and model building, I separated the dataset into two sections. The first part, the training dataset, was to be used to build the models. The second part, the verification dataset, was to be used to test the accuracy of the models after they had been built. If all data had been used to build a model, it would probably have been more accurate because it would have incorporated more data. However, there would have been no way of evaluating the model's accuracy because there would have been no independent data remaining that could be compared with the predictions of the model. If data in the training dataset were also used to test the models the results would probably have shown the models to be highly accurate. However, this accuracy would have been spurious because the data used for model-building and model-testing would have been correlated.

The verification dataset comprised forty vegetation sampling points. Ten verification points were randomly chosen from each of the four large census plots (Sewage Lagoon, Red Pine, Mature Aspen, and Peace Pipe; see Figure 2). The remaining 114 vegetation sampling points constituted the training dataset.

To incorporate information about territory location into the analysis, the territory maps developed for each species were examined. I determined for each vegetation sampling point whether it was contained within a territory of that species, i.e., whether the species was present at that point. For each point the species was scored as present, absent, or marginal (in cases where the point was so close to a territory boundary that presence or absence could not be clearly determined). This gave a qualitative measurement of whether the site was being used by that species (Present; Absent; Marginal). A point estimate of population density was also calculated as a quantitative measure of how intensely the site was being used by each species. Point densities were calculated by counting the number of territories and fractions of territories within a 0.25 ha square centered on each vegetation sampling point. Point density figures

were thus highest in areas where territories were small or closely packed. The fraction of each territory overlapping the 0.25 ha square was determined by dividing the area of the territory included inside the square into the total area of the territory. A planimeter was used to measure territory areas on the maps.

Presence/absence determinations probably had good accuracy, as the spot-mapping procedure was supplemented with song playback that indicated whether the sampling point was within a territory. Point density measurements are less reliable, as they depend on more precise knowledge of the location of territory boundaries. A possibility of error always exists in deciding where to draw the boundaries for clusters of observations. If point density measurements are accurate, however, they can potentially give more subtle information about local population distribution than can the presence/absence method.

The data collected were entered into computer spreadsheets (Microsoft Works) for analysis. Training and verification datasets were grouped separately. Each spreadsheet row held data from a different vegetation sampling point, and each column held information on a habitat variable or on the abundance of a bird species (presence/absence or point density).

Two major types of habitat-use models were developed from the information in the training dataset. The first used vegetation data to predict the probability of a species being present at a site (probability-of-presence, or POP models). These models were based on the classification of vegetation sampling points into discrete categories of bird use (Present; Absent; Marginal) as described above, and they reflected the percentage of a habitat type occupied by the bird species. The second approach used the same vegetation data to predict the density of the bird species at a site. This approach used the point estimates of density for each vegetation sampling point, as calculated above, and set up a model based on the relationships between these density figures and various habitat variables.

4.1 Probability-of-Presence (POP) Models

4.1.1. Building Rule-based and Fuzzy Models.

The initial modelling approach for each species involved a non-statistical examination of the training dataset. Habitat variables were graphed against the presence/absence information for each species, and those variables that most effectively separated unused points from used points were chosen for further examination. Interactions between the important variables were examined graphically. Rules for determining probability of presence were then assigned based on the distribution of used and unused data points on these graphs, in conjunction with my understanding of the habitat preferences of the species.

For example, a plot of Canopy Cover versus Presence/Absence (Figure 9A) indicated that Red-eyed Vireos were present at most forest-interior points but were apparently less frequent in areas with low Canopy Cover. Examination of other habitat variables indicated that few of them seemed to be related to the probability of finding Vireos; Canopy Cover appeared to have the best relationship, although it was not extremely strong. A few other variables such as Deciduous BA and Large Tree Density also had some relation to the Presence/Absence value. These were plotted against Canopy Cover in bivariate scatter plots showing occupied and unoccupied points separately (e.g., Figure 9B). These graphs suggested that the new variables were working because they were strongly correlated with Canopy Cover, and did not give useful information beyond that correlation. Therefore Canopy Cover was the only variable used in the model. The graph suggested a cutoff at 40% canopy cover, with about 70% of the closed-canopy points being occupied and about 15% of the open-canopy points being occupied. The rule was written:

70% probability of presence if Canopy Cover > 40%

15% probability of presence if Canopy Cover < 40%

Most species had more complex rules than this, involving several habitat variables. (The rule-based models for each species are summarized in Tables 3-10.) In each case several habitat types were defined based on vegetation characteristics and the probability that the species would be present was assigned for each habitat type.

Probability of presence was usually based on the percentage of points in that habitat type that were occupied. However, a good deal of personal judgement also entered into the determination of probabilities and cutoff points. With a sample size of 114 points it is possible for random sampling errors to give a misleadingly high or low number of occupied points in a habitat type, and I adjusted the probability values for the model where the data appeared potentially misleading. Thus my personal understanding of a species' habitat preferences was important in building the rule-based models.

Fuzzy models were based on the rule-based model for the species, with the same predicted probabilities of presence for each habitat type, but incorporated a gradual transition between habitat categories rather than a sharp discontinuity. For simplicity, the relationship between confidence levels and the relevant habitat variables was specified to be linear at intermediate values of the habitat variable and constant (0.0 or 1.0) at more extreme values. The confidence-level functions thus had a plateau-slope-plateau shape (Figure 10A). Graphs showing the relation between predicted probability of presence and the relevant habitat variables for each fuzzy model are given in Tables 3-10.

4.1.2. Building Logistic Regression Models.

After the rule-based model for a species had been constructed, logistic regression models were built from the training dataset, using the statistical program Statistix 3.1. The presence or absence of the bird species was used as the categorical dependent variable (scored as 0 if the species were absent at a point, 0.5 if the site was marginal, and 1 if it were present) and one or more habitat variables were used as predictor variables. Variables were initially selected by running single-variable logistic regressions against presence/absence and choosing those variables that had the best fit to the data, based on the deviance values and model p-values produced by the program. (This set of variables was usually similar to the set of variables that had been most useful in the rule-based models, as would be expected.) Curvilinear models were also run for each variable, including the variable and its square as terms in the model. Coefficient p-values of the quadratic terms were used to determine whether a significant nonlinear effect influenced the regression. A number of models was then built using

these variables, either singly or in combination, and several of these models that appeared to fit the data well were used for analysis. The models built by the program were formulae that incorporated each habitat variable, its estimated coefficient and an intercept term. The output L of each formula was then transformed by $e^L/(1+e^L)$ to yield the predicted probability of presence. If values for each habitat variable at a site were entered into the formula, a predicted value for the probability that the species was present at that site could be calculated.

4.1.3. Evaluating Probability-of-Presence Models.

When the habitat parameters in the verification dataset were entered into the models, both the rule-based and logistic regression models produced a predicted probability of the species being present for each of the forty verification points. Every verification point also had a presence/absence value for that species, indicating whether the species had actually been present there during the summer of 1991. I wrote a computer program to compare the predicted probabilities to the pattern of presence and absence that had actually occurred and determine whether a model was consistent with the observed results.

The analysis program performed a categorical statistical test designed to indicate the accuracy of the model in distinguishing used from unused habitat. The test, following Capen et al. (1986), expands upon the "proportional chance criterion" of Morrison (1969) to yield a p-value for hypothesis testing. The program sorted the verification points and their associated presence/absence data from the lowest to the highest predicted probability of presence given by the model. The verification dataset was then divided into two sections (sites predicted to be occupied and sites predicted to be unoccupied), with the cutpoint set to make the rates of correct classification as equal as possible between the two sections. Marginal points on a territory boundary were weighted as one half of a correct classification for each point. The percentage of points in the entire verification dataset that were correctly classed as present or absent was recorded. A random null model was built without regard to habitat, setting the probability of presence at all points equal to the proportion of all verification points that were actually occupied. The p-value of this test was equal to the probability of the

random null model achieving a rate of correct classification for the entire verification dataset that was greater than or equal to the percentage of points correctly classified by the habitat model. The p-value was determined by combining the binomial distributions of correct-classification rates for the predicted-occupied and predicted-unoccupied sections of the dataset. Low p-values for this test (< 0.05) indicated that the habitat model was significantly more successful than the random null model in discriminating between used and unused sites. Tables 3-10 summarize the probability-of-presence models for each species and the statistical test results for each model.

4.2 Population Density Models

4.2.1 Building Population Density Models.

The population density models were built using the multiple regression section of Statistix 3.1. Point density values from the training dataset were used as the dependent variable for the regressions. All habitat variables that appeared potentially useful were tested by running single-variable regressions against point density. Curvilinear regression models that incorporated the squares of these variables were also constructed. I used the R^2 statistic, adjusted for the number of explanatory variables, to evaluate regression models. This statistic indicated how much of the observed variation in point density was explained by the model. The coefficient p-values of quadratic terms were also used to decide whether they should be included. Variables that were effective by these criteria were included in multiple regression models. A number of models was built with various combinations of variables in an attempt to maximize the adjusted R^2 values. Several models were selected for analysis, either because they fit the training dataset particularly well or because the variables seemed likely to be biologically meaningful.

4.2.2. Evaluating Population Density Models.

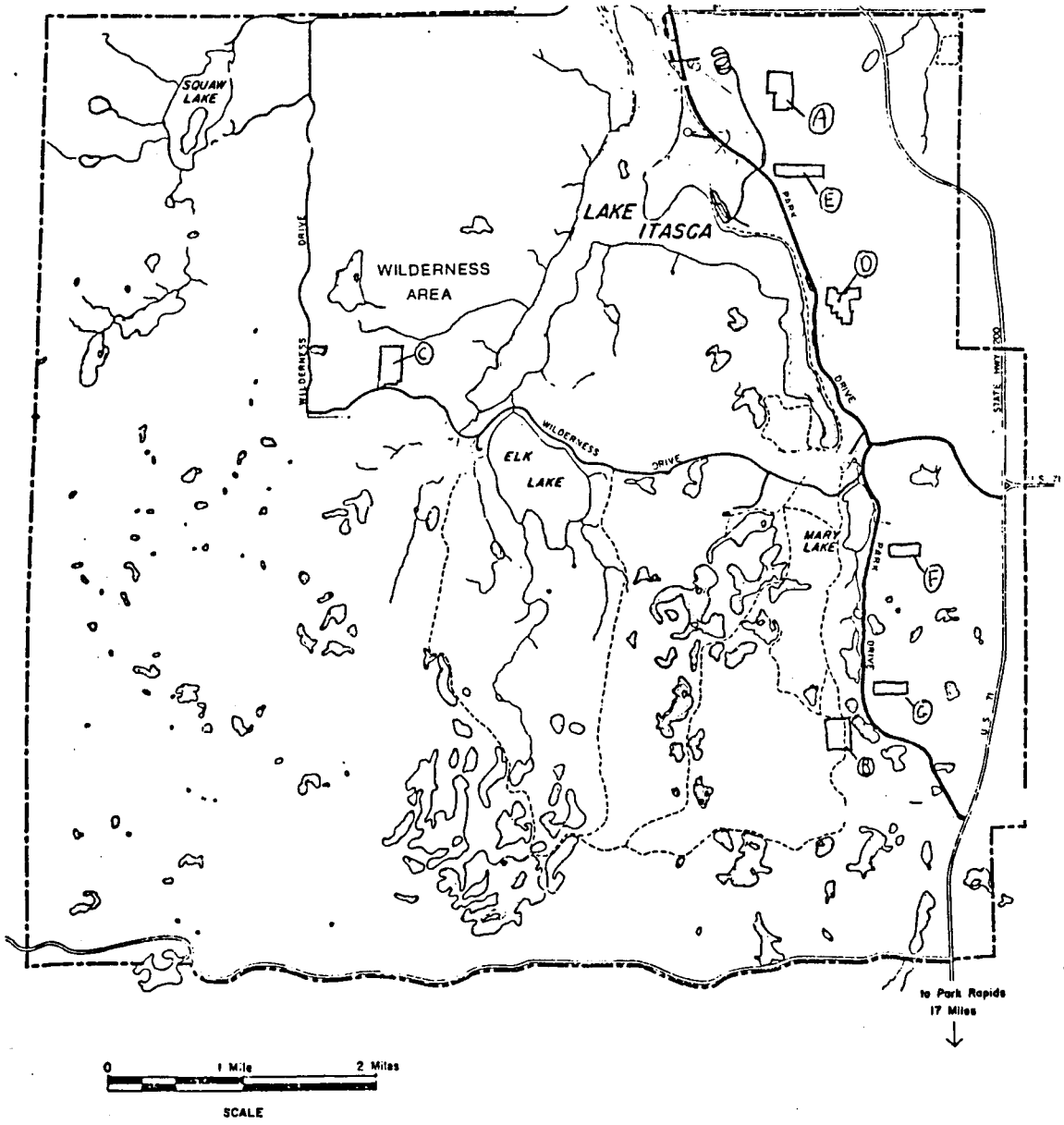
The most straightforward statistical method for comparing predicted point densities with actual point densities was simple linear regression. The actual point densities from the verification dataset were used as the dependent variable, and the predicted point densities were used as the single explanatory variable. A model that

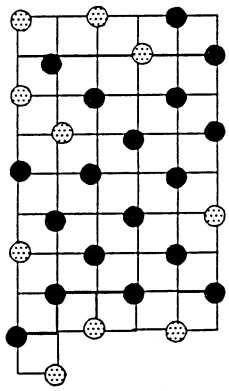
accurately predicted density would be expected to have a positive slope with a variable coefficient of 1 and a y-intercept of zero. To test whether the model's predictions were positively related to the observed data, I examined the p-value of the variable coefficient, which indicated the probability of the null hypothesis that the slope was equal to zero. Low p-values indicated a strong relationship between predictions and observed data.

Tables 11-18 summarize the regression models for each species, and also give the adjusted R^2 values from the training dataset and variable coefficient p-values for tests against the verification dataset. The models are also ranked from most to least accurate according to each of these measures of model accuracy. The few models that had a negative slope on the regression of predicted density against observed density (extremely poor predictive value) are indicated by a minus sign next to the p-value.

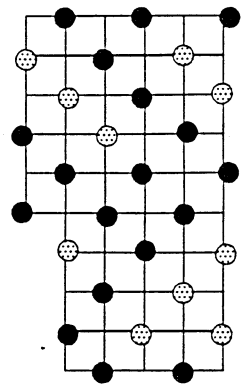
Figure 1. Map of Plot Locations in Itasca Park

A. Sewage Lagoon Plot
B. Mature Aspen Plot
C. Red Pine Plot
D. Peace Pipe Plot
E. Checkpoint Plot
F. Mary Lake Plot
G. Arco Lake Plot

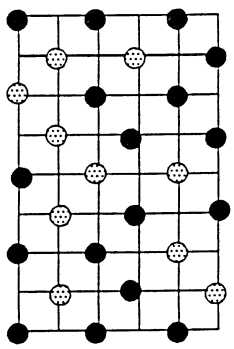




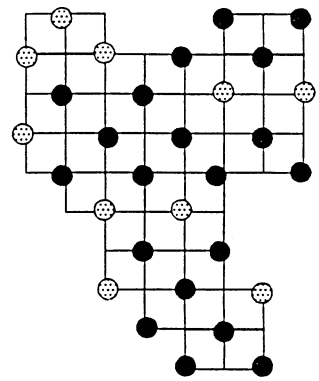
Red Pine Plot



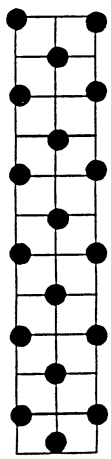
Sewage Lagoon Plot



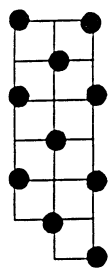
Mature Aspen Plot



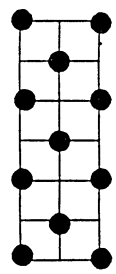
Peace Pipe Plot



Checkpoint Plot



Mary Lake Plot



Arco Lake Plot

100 meters

Vegetation Sampling Point
(Training Dataset)

Vegetation Sampling Point
(Verification Dataset)

Figure 2. Census Plot Maps and Vegetation Sampling Points

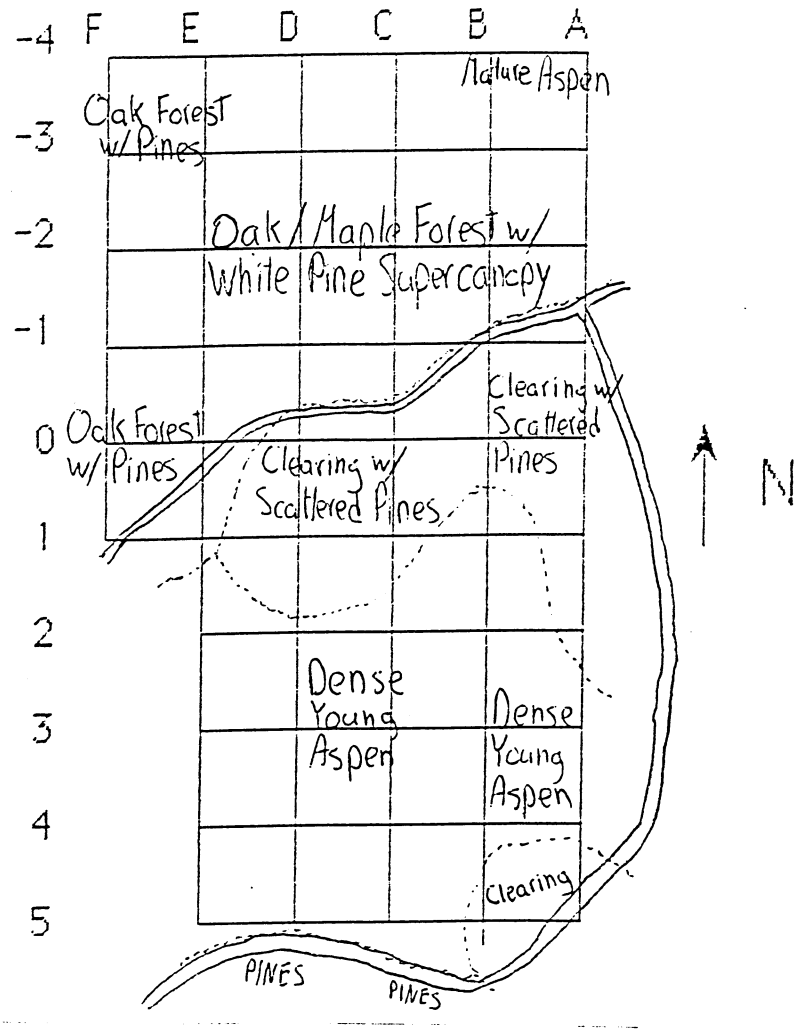


Figure 3. Vegetation Map of Sewage Lagoon Plot.

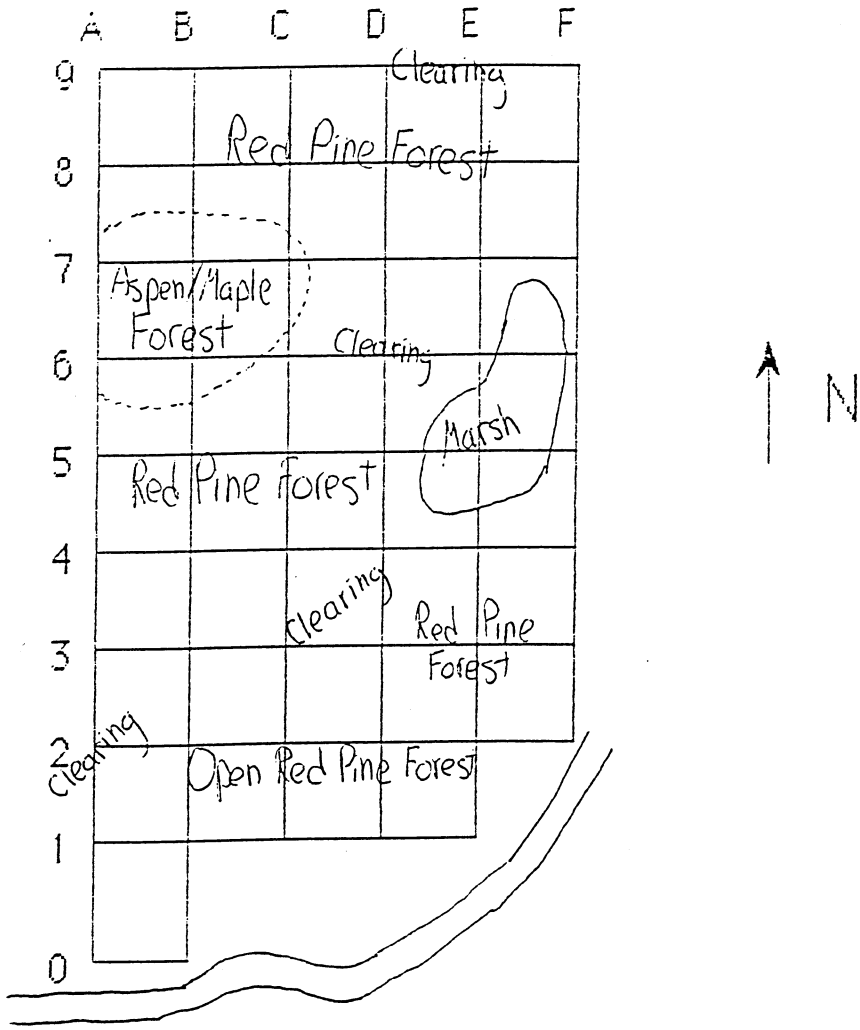


Figure 4. Vegetation Map of Red Pine Plot.

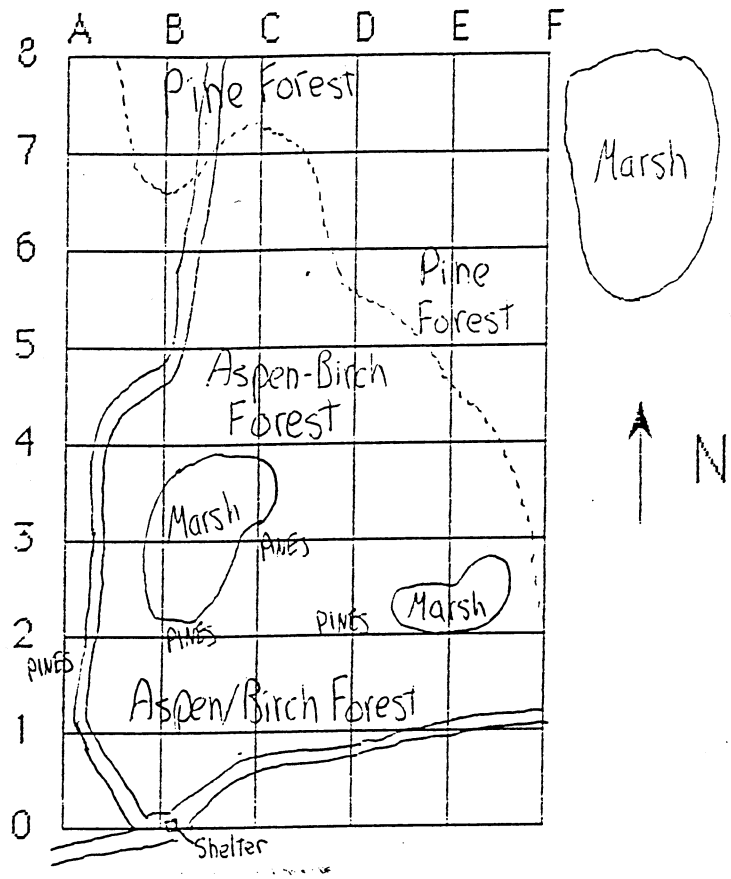


Figure 5. Vegetation Map of Mature Aspen Plot.

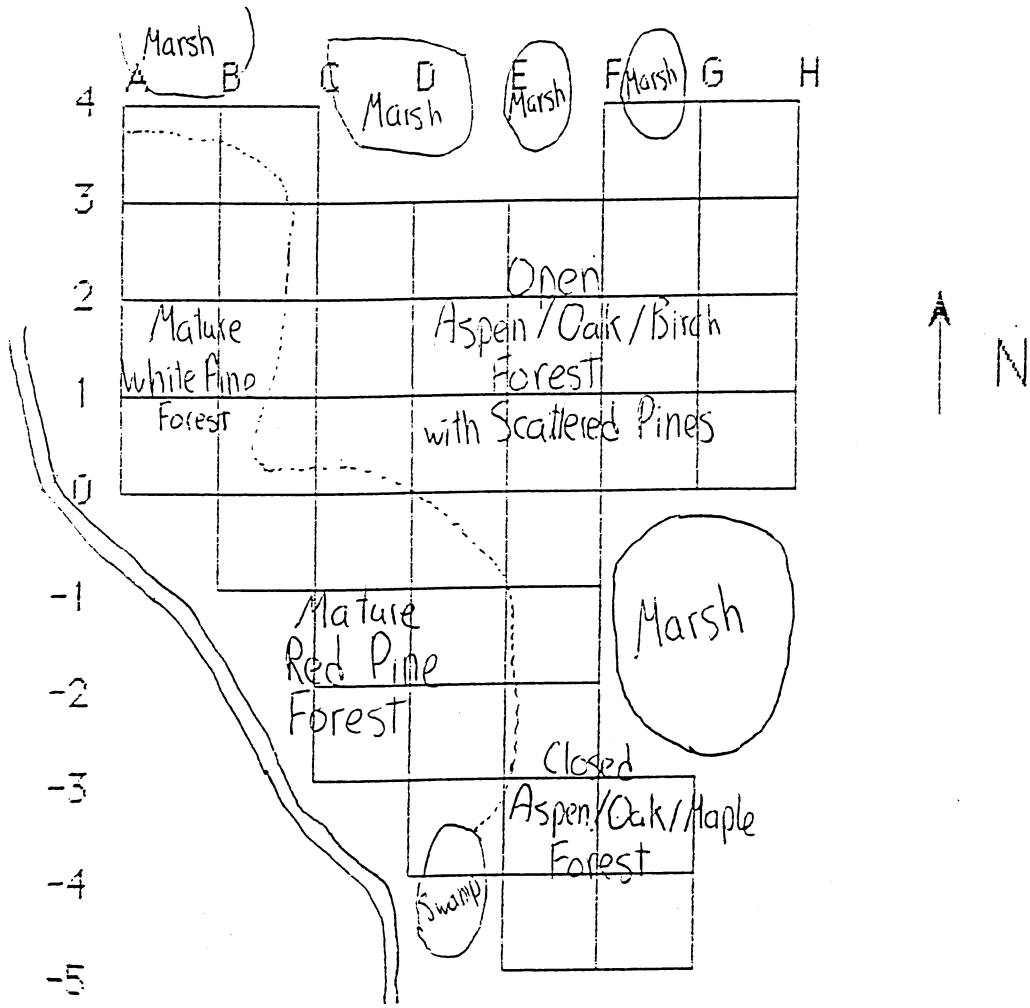


Figure 6. Vegetation Map of Peace Pipe Plot.

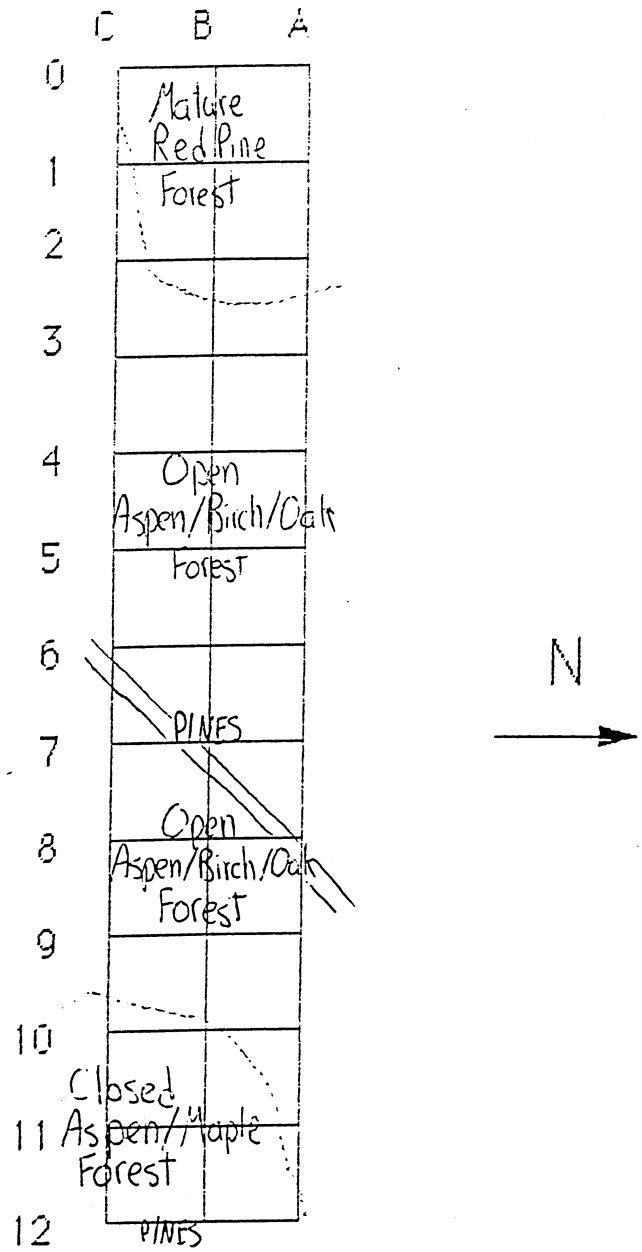


Figure 7. Vegetation Map of Checkpoint Plot.

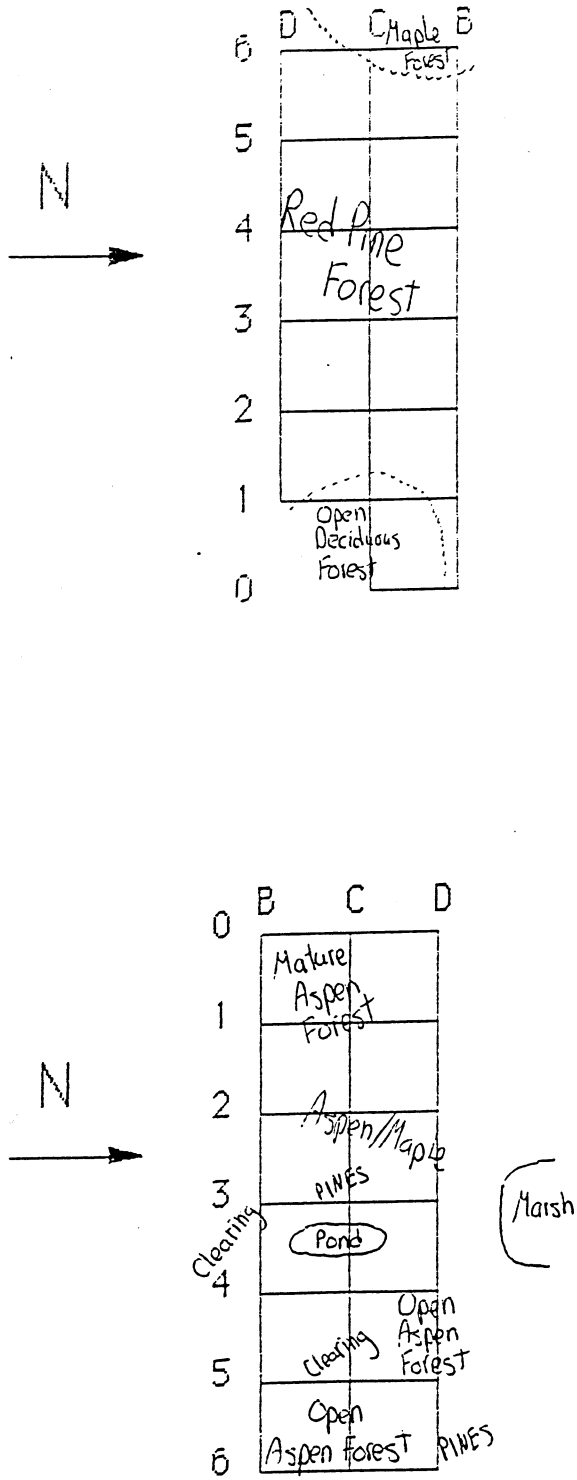
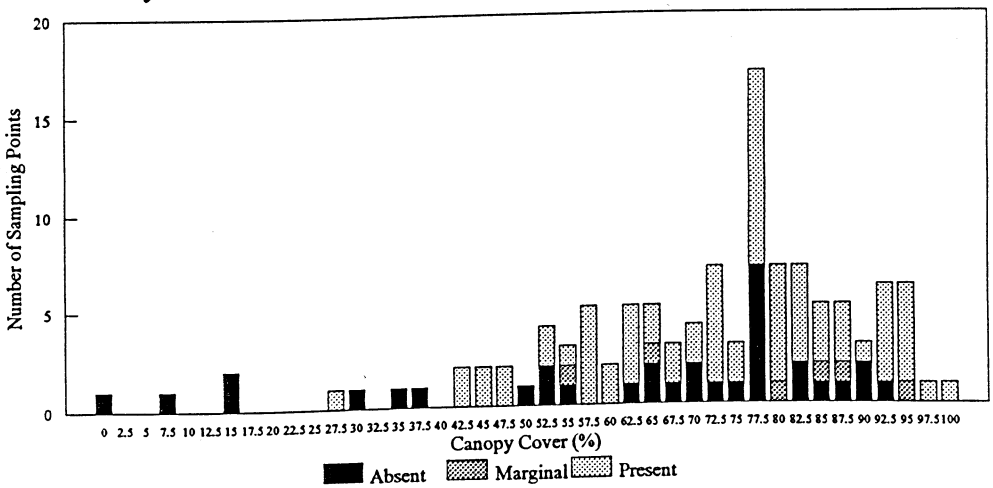


Figure 8. Vegetation Map of Mary Lake and Arco Lake Plots.

Figure 9. Examples of Habitat Data (Red-eyed Vireo).

A. Red-eyed Vireo Occurence vs. Canopy Cover.



B. Canopy Cover vs. Deciduous Basal Area, Showing Vireo Occurrence.

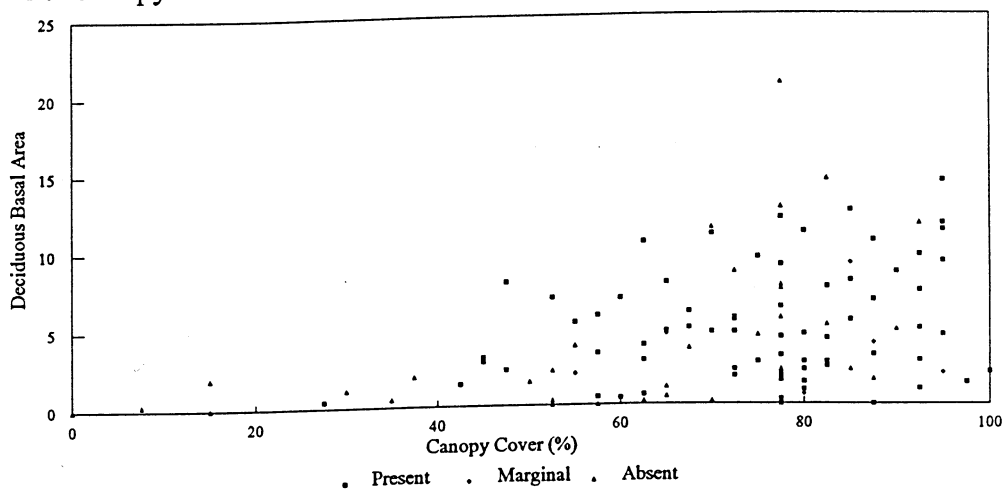
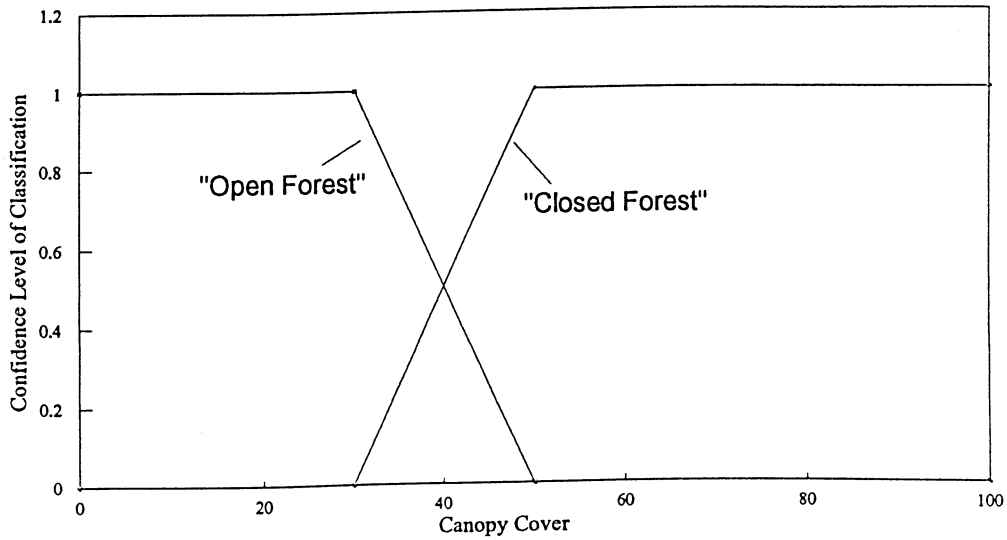


Figure 10. Fuzzy Model Confidence Levels and Model Predictions (Red-eyed Vireo).

A. Confidence Level Functions for Red-eyed Vireo Fuzzy Model.



B. Predictions of Vireo Occurrence from Several Modelling Methods.

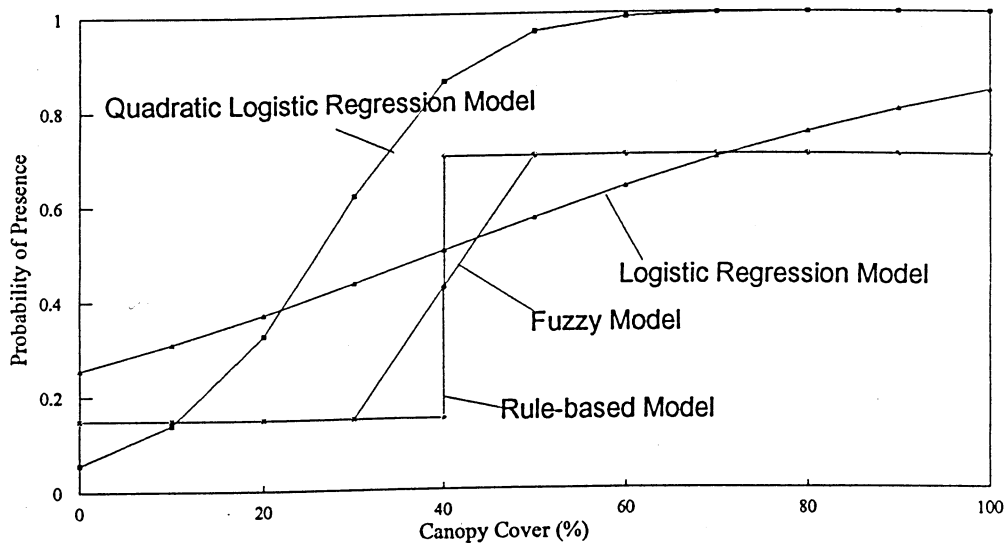


Figure 11. Sample Expert System Session.

Context:
You are in an upland forest in Itasca State Park during June. This system will indicate the probability that the spot where you are standing is included in the territory of a specified species of forest songbird.
Question:
What species of bird are you interested in?

1) Red-eyed Vireo
2) Common Yellowthroat
3) Chestnut-sided Warbler
4) Pine Warbler
5) Black-throated Green Warbler
6) Ovenbird
7) American Redstart
8) Least Flycatcher

Press the number of the answer or W [Why].
> W [Why]
Different species of birds have different habitat requirements.
> 3 [Chestnut-sided Warbler]
Question:
What is the total basal area of deciduous trees within a 0.1 acre circle?

1) Less than 6 square feet
2) Greater than or equal to 6 square feet

Press the number of the answer or W [Why].
> W [Why]
Deciduous basal area indicates whether a site is dominated by mature deciduous trees.
> W [Why]
Chestnut-sided Warblers occur in early successional forests or open pine forests, which have open canopies and few large deciduous trees. They are rare in mature, closed-canopy deciduous forest.
> 1 [Less than 6 square feet]
Question:
What is the local canopy cover within a 0.1 acre circle?

1) Greater than 70%
2) Between 51 and 70%
3) Between 41% and 50%
4) Between 26% and 40%
5) Less than or equal to 25%

Press the number of the answer or W [Why].
> W [Why]
A bird may use canopy cover as a rule of thumb to indicate habitat quality. Canopy cover is related to foliage density, insect availability, cover from predators and weather, and density of ground cover.
> W [Why]
Chestnut-sided Warblers occur in early successional forests or open pine forests, which have open canopies and few large deciduous trees. They are rare in mature, closed-canopy deciduous forest.
> 4 [Between 26% and 40%]
Decision:
Moderate/high quality: 40% chance of Chestnut-sided Warbler being present.

There are no more valid decisions.

Table 1. Habitat Characteristics of Breeding Bird Census Plots

Census Plot	Mature Pine Forest	Mature Aspen/Birch Forest	Mature Oak/Maple Forest	Young Aspen Forest	Upland Canopy Openings	Open Marshes or Bogs
Sewage Lagoon (10.25 ha)	Scattered Trees (10%)	Patchy (15%)	Patchy (15%)	Extensive (35%)	Large Gaps (25%)	None
Red Pine (10.0 ha)	Extensive (60%)	Patchy (10%)	Scattered Trees (5%)	None	Large Gaps (15%)	One (10%)
Mature Aspen (10.0 ha)	Patchy (20%)	Patchy (30%)	Patchy (15%)	None	Small Gaps Only (10%)	Two (10%)
Peace Pipe (10.0 ha)	Patchy (20%)	Patchy (30%)	Patchy (20%)	None	Large Gaps (30%)	Bordering Plot (5%)
Checkpoint (6 ha)	Patchy (15%)	Extensive (40%)	Patchy (15%)	None	Large Gaps (10%)	None
Mary Lake (2.75 ha)	Extensive (80%)	Scattered Trees (5%)	Scattered Trees (5%)	None	Large Gap (10%)	None
Arco Lake (3.0 ha)	Scattered Trees (5%)	Extensive (60%)	Scattered Trees (10%)	None	Large Gaps (25%)	None

Table 2. Description and Observed Ranges of Habitat Variables

Habitat Variable	Units	Range of Observed Values for Variable (Percentile)			
		Low	5th	95th	High
Pine Proximity	Categorical Variable: 0 (>25m); 1 (12-25m); 2 (<12m)	0	0	2	2
Marsh Proximity	Categorical Variable: 0 (>25m); 1 (12-25m); 2 (<12m)	0	0	1	2
Pine Basal Area (BA)	Square feet/plot*	0	0	21.3	35.1
Red Pine BA	Square feet/plot	0	0	17.2	24.5
White Pine BA	Square feet/plot	0	0	15.1	32
Spruce BA	Square feet/plot	0	0	0.2	1.8
Fir BA	Square feet/plot	0	0	1.3	3.6
Aspen BA	Square feet/plot	0	0	8.5	19.9
Birch BA	Square feet/plot	0	0	3.5	5.9
Oak BA	Square feet/plot	0	0	3.4	13.2
Maple BA	Square feet/plot	0	0	3	5.2
Ironwood BA	Square feet/plot	0	0	0.6	1.6
Deciduous BA	Square feet/plot	0	0.2	12.4	20.6
Coniferous BA	Square feet/plot	0	0	21.3	35.1
Large Tree Density	Trees/acre (> 9" dbh)	0	0	130	170
Small Tree Density	Trees/acre (3-9" dbh)	0	20	220	430
Shrub Density	Stems/acre (< 3" dbh)	500	1300	13200	16800
Ground Cover	Percent (< 1 m height)	7.5	15	90	100
Subcanopy Cover	Percent (2-10 m height	0	12.5	85	92.5
Canopy Cover	Percent (2+ m height)	0	35	95	100
Deciduous Subcanopy	Percent (2-10 m height	0	12.5	82.5	92.5
Deciduous Canopy	Percent (2+ m height)	0	12.5	92.5	95
Pine Canopy	Percent (2+ m height)	0	0	40	55
Ironwood Canopy	Percent (2+ m height)	0	0	50	65
Maple Canopy	Percent (2+ m height)	0	0	75	92.5
Oak Canopy	Percent (2+ m height)	0	0	30	52.5

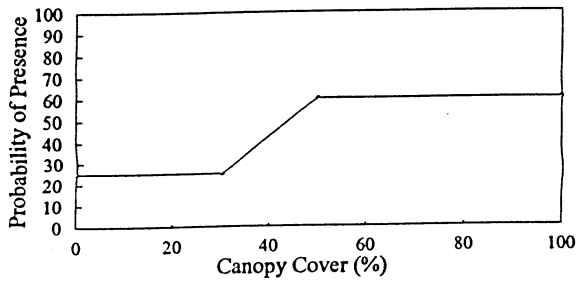
* Vegetation sampling plots are 0.1 acres (0.04 ha) in area.

Table 3. Ovenbird Probability-of-Presence Models

Rule-based Model

IF CanopyCover <= 40%	THEN 25% Probability of presence	(Open Forest)
IF CanopyCover > 40%	THEN 60% Probability of presence	(Closed Forest)

Fuzzy Model: Probability of Presence



Logistic Regression Models

Predicted probability of presence is equal to (e^L/(1 + e^L))

Model A L = -0.07 + 0.07(ConiferousBA)
Model B L = -0.88 + 0.018(CanopyCover)
Model C L = -1.26 + 0.017(CanopyCover) + 0.066(ConiferousBA)
Model D L = -0.31 + 0.038(DeciduousBA) + 0.079(ConiferousBA)
Model E L = -4.33 + 0.063(CanopyCover) - 0.000373(CanopyCover)^2
+ 0.000566(ShrubDensity) - 0.0000000344(ShrubDensity)^2
+ 0.0768(ConiferousBA)

Tests of Models Against Verification Dataset (n=40)

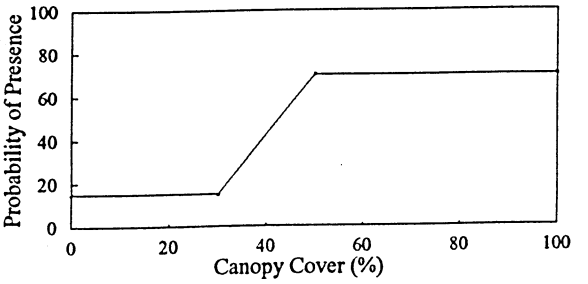
	p-value
Rule-based Model	0.21
Fuzzy Model	0.09
Logistic Model A	0.07
Logistic Model B	0.26
Logistic Model C	0.26
Logistic Model D	0.43
Logistic Model E	0.10

Table 4. Red-eyed Vireo Probability-of-Presence Models

Rule-based Model

IF CanopyCover <= 40%	THEN 15% Probability of presence	(Open Forest)
IF CanopyCover > 40%	THEN 70% Probability of presence	(Closed Forest)

Fuzzy Model: Probability of Presence



Logistic Regression Models

Predicted probability of presence is equal to (e^L/(1 + e^L))

Model A L = -2.81 + 0.093(CanopyCover) - 0.00056(CanopyCover)^2
Model B L = -1.07 + 0.027(CanopyCover)
Model C L = -1.23 + 0.070(LargeTreeDensity) - 0.000465(LargeTreeDensity)^2

Tests of Models Against Verification Dataset (n=40)

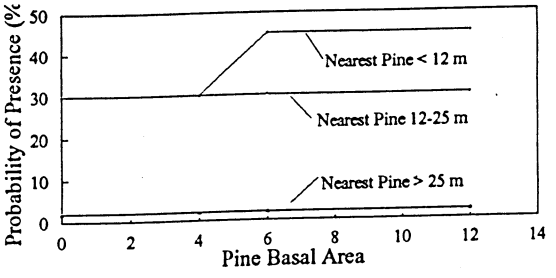
	p-value
Rule-based Model	0.54
Fuzzy Model	0.36
Logistic Model A	0.36
Logistic Model B	0.36
Logistic Model C	0.20

Table 5. Pine Warbler Probability-of-Presence Models

Rule-based Model

IF nearest pine > 25 m away	THEN 2% Probability of presence	(Not Pine Forest)
IF nearest pine is 12-25 m away	THEN 30% Probability of presence	(Scattered Pine Forest)
IF 0 sq.ft < Pine BA < 5 sq.ft	THEN 30% Probability of presence	(Young Pine Forest)
IF Pine BA >= 5 sq. ft	THEN 45% Probability of presence	(Mature Pine Forest)

Fuzzy Model: Probability of Presence



Logistic Regression Models

Predicted probability of presence is equal to (e^L/(1 + e^L))

Model A L = -2.87 + 1.66(PineProximity) - 0.21(PineBA) + 0.0082(PineBA)^2

Model B L = -2.75 + 1.23(PineProximity) - 0.139(WhitePineBA) + 0.011(WhitePineBA)^2 + 0.0052(RedPineBA) - 0.00173(RedPineBA)^2

Model C L = -1.46 + 0.063(PineBA)

Tests of Models Against Verification Dataset (n=40)

	p-value
Rule-based Model	0.019
Fuzzy Model	0.008
Logistic Model A	0.019
Logistic Model B	0.56
Logistic Model C	0.008

Table 6. Common Yellowthroat Probability-of-Presence Models

Rule-based Model

IF no marsh within 25 m AND CanopyCover> 25%	THEN 2% Probability of presence	(Upland Forest)
IF a marsh is 12-25 m away AND CanopyCover > 25%	THEN 20% Probability of presence	(Marsh Border Forest)
IF no marsh within 12 m AND CanopyCover <= 25%	THEN 50% Probability of presence	(Upland Clearing)
IF a marsh is within 12 m	THEN 80% Probability of presence	(Marsh)

Logistic Regression Models

Predicted probability of presence is equal to (e^L/(1 + e^L))

Model A L = 13.28 - 0.9(CanopyCover) + 0.0001(CanopyCover)^2 + 60.66(MarshProximity) - 17.95(MarshProximity)^2

Model B L = -4.081 + 3.24(MarshProximity)

Model C L = 2.268 - 0.091(CanopyCover)

Model D L = 13.36 - 0.89(CanopyCover) + 46.62(MarshProximity)

Tests of Models Against Verification Dataset (n=40)

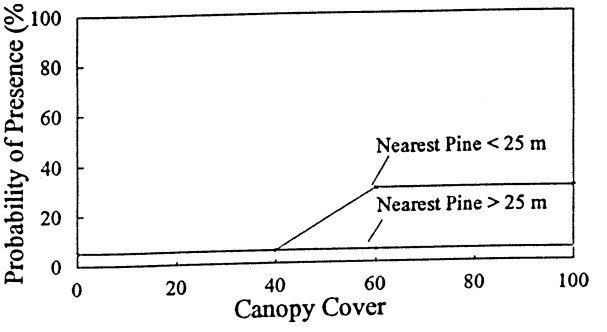
	p-value
Rule-based Model	0.009
Logistic Model A	0.009
Logistic Model B	0.009
Logistic Model C	0.38
Logistic Model D	0.009

Table 7. Black-throated Green Warbler Probability-of-Presence Models

Rule-based Model

IF CanopyCover <= 50%	THEN 5% Probability of presence	(Open Forest)
IF CanopyCover > 50% AND nearest pine > 25 m away	THEN 5% Probability of presence	(Closed Deciduous Forest)
IF CanopyCover > 50% AND nearest pine <= 25 m away	THEN 30% Probability of presence	(Closed Forest with Pine)

Fuzzy Model: Probability of Presence



Logistic Regression Models

Predicted probability of presence is equal to (e^L/(1 + e^L))

Model A L = -4.77 + 0.73(PineProximity) + 0.032(CanopyCover)

Model B L = -4.77 + 0.072(PineProximity) + 0.0012(PineBA) + 0.032(CanopyCover)

Model C L = -7.7 + 0.72(PineProximity) + 0.12(CanopyCover) - 0.00061(CanopyCover)^2

Model D L = -2.40 + 0.723(PineProximity)

Model E L = -3.68 + 0.32(CanopyCover)

Tests of Models Against Verification Dataset (n=40)

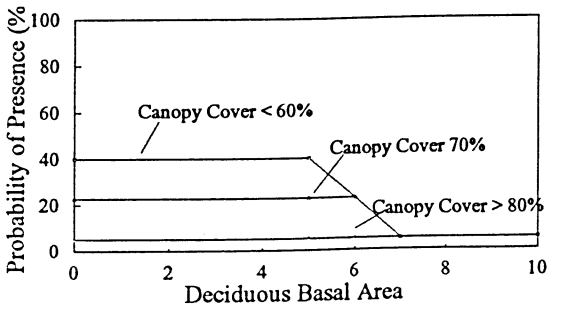
	p-value
Rule-based Model	0.18
Fuzzy Model	0.53
Logistic Model A	0.21
Logistic Model B	0.21
Logistic Model C	0.21
Logistic Model D	0.39
Logistic Model E	0.21

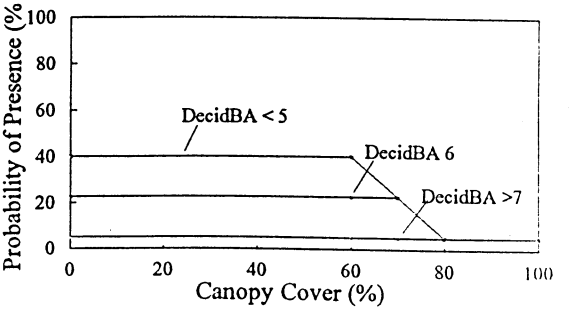
Table 8. Chestnut-sided Warbler Probability-of-Presence Models

Rule-based Model

IF DeciduousBA <= 6.0 sq. ft AND CanopyCover < 70%	THEN 40% Probability of presence	(Not Mature Deciduous Forest)
IF Deciduous BA > 6.0 sq. ft OR CanopyCover >= 70%	THEN 5% Probability of presence	(Mature Deciduous Forest)

Fuzzy Model: Probability of Presence





Logistic Regression Models

Predicted probability of presence is equal to (e^L/(1 + e^L))

Model A	L = 0.91 - 0.045(CanopyCover)
Model B	L = -2.13 + 0.113(CanopyCover) - 0.0016(CanopyCover)^2
Model C	L = -1.18 - 0.22(DecidBA)
Model D	L = 0.17 - 0.052(CanopyCover) + 0.00017(ShrubDensity)
Model E	L = -2.21 + 0.09(CanopyCover) - 0.0015(CanopyCover)^2 + 0.000115(ShrubDensity)

Tests of Models Against Verification Dataset (n=40)

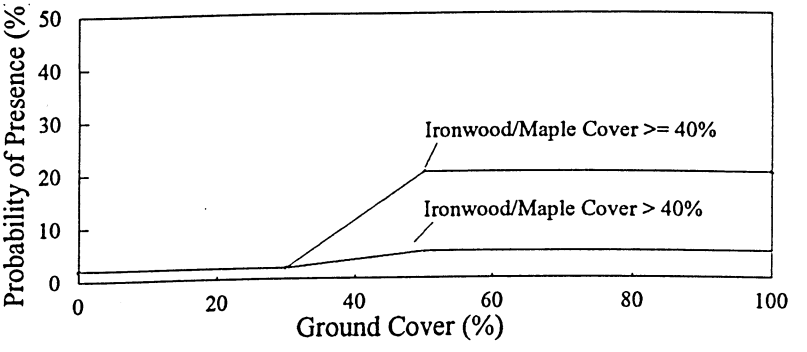
	p-value
Rule-based Model	0.0005
Fuzzy Model	0.00009
Logistic Model A	0.008
Logistic Model B	0.008
Logistic Model C	0.20
Logistic Model D	0.001
Logistic Model E	0.008

Table 9. American Redstart Probability-of-Presence Models

Rule-based Model

IF GroundCover >= 40% AND Ironwood/MapleCover < 40%	THEN 20% Probability of presence	(Open Forest)
IF GroundCover >= 40% AND Ironwood/MapleCover >= 40	THEN 5% Probability of presence	(Open Maple Forest)
IF GroundCover < 40%	THEN 2% Probability of presence	(Forest Interior)

Fuzzy Model: Probability of Presence



Logistic Regression Models

Predicted probability of presence is equal to (e^L/(1 + e^L))

Model A	L = -4.59 + 0.042(GroundCover)
Model B	L = -7.03 + 0.126(GroundCover) - 0.00066(GroundCover)^2
Model C	L = -7.77 + 0.029(CanopyCover) + 0.062(GroundCover)
Model D	L = -5.46 + 0.000124(ShrubDensity) + 0.043(GroundCover)
Model E	L = -3.12 + 0.0283(GroundCover) - 0.0256(Ironwood/MapleCover)

Tests of Models Against Verification Dataset (n=40)

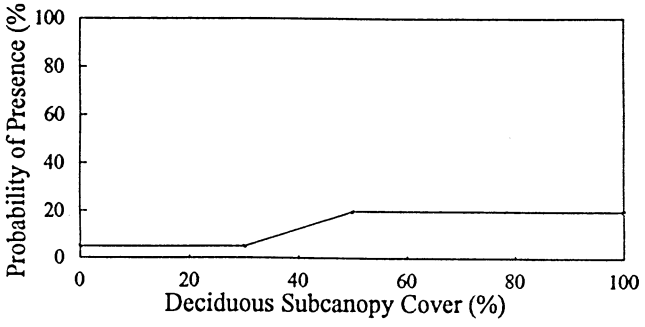
	p-value
Rule-based Model	0.08
Fuzzy Model	0.14
Logistic Model A	0.12
Logistic Model B	0.12
Logistic Model C	0.12
Logistic Model D	0.37
Logistic Model E	0.11

Table 10. Least Flycatcher Probability-of-Presence Models

Rule-based Model

IF DecidSubcanopyCover > 40%	THEN 20% Probability of presence	(Closed Subcanopy)
IF DecidSubcanopyCover <= 40%	THEN 5% Probability of presence	(Open Subcanopy)

Fuzzy Model: Probability of Presence



Logistic Regression Models

Predicted probability of presence is equal to (e^L/(1 + e^L))

Model A	L = -3.29 + 0.0212(CanopyCover)
Model B	L = -53.41 + 1.41(CanopyCover) - 0.0094(CanopyCover)^2
Model C	L = -2.62 + 0.161(DecidBA)
Model D	L = -3.08 + 0.0071(CanopyCover) + 0.151(DeciduousBA)

Tests of Models Against Verification Dataset (n=40)

	p-value
Rule-based Model	0.95
Fuzzy Model	0.98
Logistic Model A	0.97
Logistic Model B	0.88
Logistic Model C	0.99
Logistic Model D	0.99

Table 11. Population Density Models: Ovenbird

Predicted density (Territories/0.25 ha) is equal to D.

Model I	D = 0.166 + 0.00694(WhitePineBA) + 0.0177(RedPineBA) + 0.00206(CanopyCover)
Model II	D = 0.248 + 0.0158(RedPineBA) + 0.00732(WhitePineBA) + 0.0018(LargeTreeDensity)
Model III	D = 0.242 + 0.0118(PineBA) + 0.00118(LargeTreeBA)
Model IV	D = 0.31 + 0.00784(WhitePineBA) + 0.0175(RedPineBA)
Model V	D = 0.311+ 0.0132(PineBA)
Model VI	D = 0.294 + 0.0192(PineProximity) + 0.0119(PineBA)
Model VII	D = 0.274 + 0.00191(LargeTreeDensity)
Model VIII	D = 0.091 + 0.00155(LargeTreeDensity) + 0.00683*CanopyCover - 0.0000513(CanopyCover)^2
Model IX	D = 0.237 + 0.000682CanopyCover) + 0.00173(LargeTreeDensity)
Model X	D = 0.056 + 0.01(CanopyCover) - 0.0000678(CanopyCover)^2
Model XI	D = 0.252 + 0.00211(CanopyCover)
Model XII	D = 0.493 - 0.00176(GroundCover)

Tests of Ovenbird Density Models

	Training Dataset		Verification Dataset		
Model	AdjR^2	Model Rank	p-value	Model Rank	Variables Used
I	0.213	1	0.0086	7	RedPineBA WhitePineBA CanopyCover
II	0.207	2	0.0014	5	RedPineBA WhitePineBA LargeTreeDensity
III	0.196	3	0.0005	2	PineBA LargeTreeDensity
IV	0.194	4	0.041	12	RedPineBA WhitePineBA
V	0.177	5	0.028	11	PineBA
VI	0.172	6	0.024	10	PineProximity PineBA
VII	0.063	7	0.0006	4	LargeTreeDensity
VIII	0.062	8	0.0006	3	LargeTreeDensity CanopyCover CanopyCover^2
IX	0.057	9	0.0003	1	CanopyCover LargeTreeDensity
X	0.035	10	0.01	8	CanopyCover CanopyCover^2
XI	0.019	11	0.0019	6	CanopyCover
XII	0.017	12	0.013	9	GroundCover

Table 12. Population Density Models: Red-eyed Vireo

Predicted density (Territories/0.25 ha) is equal to D.

Model I D = 0.311 + 0.00913(LargeTreeDensity)
- 0.0000586(LargeTreeDensity) ^2
- 0.000864(SmallTreeDensity)

Model II D = 0.23 + 0.00926(LargeTreeDensity)
- 0.0000594(LargeTreeDensity)

Model III D = 0.763 - 0.00299(GroundCover) - 0.00105(SmallTreeDensity)

Model IV D = 0.376 + 0.0035(CanopyCover - 0.00105(SmallTreeDensity)

Model V D = 0.656 - 0.00276(GroundCover)

Model VI D = 0.533 + 0.00128(CanopyCover) - 0.00212(GroundCover)

Model VII D = 0.322 + 0.00267(CanopyCover)

Model VIII D = 0.595 - 0.000926(SmallTreeDensity)

Model IX D = 0.452 + 0.00178(Ostrya/Maple Cover)

Model X D = 0.466 + 0.000694(LargeTreeDensity)

Tests of Red-eyed Vireo Density Models

	Training Dataset		Verification Dataset		
Model	AdjR^2	Model Rank	p-value	Model Rank	Variables Used
I	0.124	1	0.90	6	LargeTreeDensity LargeTreeDensity^2 SmallTreeDensity
II	0.108	2	0.59 (-)	10	LargeTreeDensity LargeTreeDensity^2
III	0.064	3	0.59	5	GroundCover SmallTreeDensity
IV	0.061	4	0.46	4	CanopyCover SmallTreeDensity
V	0.037	5	0.80 (-)	9	GroundCover
VI	0.033	6	0.83 (-)	8	CanopyCover GroundCover
VII	0.023	7	0.93 (-)	7	CanopyCover
VIII	0.019	8	0.33	3	SmallTreeDensity
IX	0.017	9	0.21	2	Ostrya/MapleCover
X	-0.002	10	0.071	1	LargeTreeDensity

Table 13. Population Density Models: Pine Warbler

Predicted density (Territories/0.25 ha) is equal to D.

Model I	D = 0.034 + 0.0129(RedPineBA) + 0.113(PineProximity)
Model II	D = 0.036 + 0.0133(RedPineBA) + 0.00189(WhitePineBA) + 0.1069(PineProximity)
Model III	D = 0.035 + 0.00805(PineBA) + 0.110(PineProximity)
Model IV	D = 0.024 + 0.159(PineProximity)
Model V	D = 0.129 + 0.021(RedPineBA) + 0.00887(WhitePineBA)
Model VI	D = 0.154 + 0.0211(RedPineBA)
Model VII	D = 0.131+ 0.0156(PineBA)
Model VIII	D = 0.152 + 0.00948(PineCover)
Model IX	D = 0.211 + 0.00912(WhitePineBA)
Model X	D = 0.304 - 0.000947(CanopyCover)

Tests of Pine Warbler Density Models

	Training Dataset		Verification Dataset		
Model	AdjR^2	Model Rank	p-value	Model Rank	Variables Used
I	0.262	1	0.011	5	RedPineBA PineProximity
II	0.256	2	0.0084	4	PineProximity RedPineBA WhitePineBA
III	0.237	3	0.0016	1	PineBA PineProximity
IV	0.215	4	0.0025	2	PineProximity
V	0.203	5	0.037	7	RedPineBA WhitePineBA
VI	0.183	6	0.21	9	RedPineBA
VII	0.182	7	0.0058	3	PineBA
VIII	0.180	8	0.13	8	PineCover
IX	0.020	9	0.012	6	WhitePineBA
X	-0.005	10	0.28	10	CanopyCover

Table 14. Population Density Models: Common Yellowthroat

Predicted density (Territories/0.25 ha) is equal to D.

Model I	D = 0.181 + 0.128(MarshProximity) - 0.00168(TreeDensity) + 0.00000346(TreeDensity)^2
Model II	D = 0.139 - 0.000354(SmallTreeDensity) - 0.000533(LargeTreeDensity) - 0.000794(CanopyCover) + 0.135(MarshProximity)
Model III	D = 0.101 - 0.000418(SmallTreeDensity) - 0.000729(LargeTreeDensity) + 0.143(MarshProximity)
Model IV	D = 0.094 - 0.000511(TreeDensity) + 0.143(MarshProximity)
Model V	D = 0.118 - 0.00145(CanopyCover) + 0.137(MarshProximity)
Model VI	D = 0.134 + 0.000168(GroundCover) - 0.138(MarshProximity) - 0.00156(CanopyCover)
Model VII	D = 0.013 + 0.156(MarshProximity)
Model VIII	D = 0.288 - 0.00274(TreeDensity) + 0.00000612(TreeDensity)^2
Model IX	D = 0.208 - 0.00247(CanopyCover)
Model X	D = 0.144 - 0.000711(TreeDensity)
Model XI	D = -0.028 + 0.00118(GroundCover)

Tests of Common Yellowthroat Density Models

	Training Dataset		Verification Dataset		
Model	AdjR^2	Model Rank	p-value	Model Rank	Variables Used
I	0.458	1	0.0094	6	MarshProximity TreeDensity TreeDensity^2
II	0.436	2	0.002	4	MarshProximity LargeTreeDensity SmallTreeDensity CanopyCover
III	0.429	3	0.0093	5	MarshProximity LargeTreeDensity SmallTreeDensity
IV	0.429	4	0.0097	7	MarshProximity TreeDensity
V	0.410	5	0.0001	1	MarshProximity CanopyCover
VI	0.406	6	0.0001	2	MarshProximity CanopyCover GroundCover
VII	0.365	7	0.0003	3	MarshProximity
VIII	0.240	8	0.92	11	TreeDensity TreeDensity^2
IX	0.153	9	0.23	9	CanopyCover
X	0.129	10	0.91	10	TreeDensity
XI	0.041	11	0.11	8	GroundCover

Table 15. Population Density Models: Black-throated Green Warbler

Predicted density (Territories/0.25 ha) is equal to D.

Model I D = 0.05 + 0.0042(RedPineBA) +0.0352(PineProximity) + 0.0812(BalsamFirBA)

Model II D = -0.07 + 0.0319(PineProximity) + 0.00624(RedPineBA) + 0.00187(CanopyCover)

Model III D = 0.11 + 0.0968(BalsamFirBA)

Model IV D = 0.06 + 0.039(PineProximity) + 0.00555(RedPineBA)

Model V D = 0.10 + 0.00834(RedPineBA)

Model VI D = 0.05 + 0.0582(PineProximity)

Model VII D = 0.19 - 0.0122(DeciduousBA)

Model VIII D = 0.09 + 0.00832(RedPineBA) + 0.00289(WhitePineBA)

Model IX D = 0.09 + 0.0059 (PineBA)

Model X D = -0.01 + 0.00195(CanopyCover)

Model XI D = 0.02 + 0.00326(LargeTreeDensity) - 0.0000194(LargeTreeDensity)^2

Tests of Black-throated Green Warbler Density Models

	Training Dataset		Verification Dataset		
Model	AdjR^2	Model Rank	p-value	Model Rank	Variables Used
I	0.103	1	0.91 (-)	9	RedPineBA BalsamFirBA PineProximity
II	0.081	2	0.26	3	PineProximity RedPineBA CanopyCover
III	0.067	3	0.73 (-)	10	BalsamFirBA
IV	0.059	4	0.92	6	PineProximity RedPineBA
V	0.049	5	0.94 (-)	8	RedPineBA
VI	0.049	6	0.81	4	PineProximity
VII	0.046	7	0.24 (-)	11	DeciduousBA
VIII	0.046	8	0.99	7	RedPineBA WhitePineBA
IX	0.043	9	0.87	5	PineBA
X	0.025	10	0.0077	1	CanopyCover
XI	0.013	11	0.13	2	LargeTreeDensity LargeTreeDensity^2

Table 16. Population Density Models: Chestnut-sided Warbler

Predicted density (Territories/0.25 ha) is equal to D.

Model I	D = 0.198 + 0.0025(CanopyCover) - 0.000048(CanopyCover)^2 - 00012(LargeTreeDensity) + 0.0000081(ShrubDensity)
Model II	D = 0.202 + 0.00445(CanopyCover) - 0.0000646(CanopyCover)^2 - 0.00123(LargeTreeDensity)
Model III	D = 0.224 - 0.0000907(CanopyCover) - 0.0000341(CanopyCover)^2 + 0.00000858(ShrubDensity)
Model IV	D = 0.214 - 0.00301(CanopyCover) + 0.00165(GroundCover)
Model V	D = 0.386 - 0.00329(CanopyCover) - 0.000998(LargeTreeDensity)
Model VI	D = 0.229 + 0.00191(CanopyCover) - 0.0000515(CanopyCover)^2
Model VII	D = 0.378 - 0.00411(CanopyCover)
Model VIII	D = 0.301 - 0.0051(LargeTreeDensity) + 0.0000225(LargeTreeDensity)^2
Model IX	D = -0.076 + 0.00314(GroundCover)
Model X	D = 0.212 - 0.00186(LargeTreeDensity)
Model XI	D = 0.021 + 0.0000110(ShrubDensity)
Model XII	D = 0.129 - 0.00902(DeciduousBA)

Tests of Chestnut-sided Warbler Density Models

	Training Dataset		Verification Dataset		
Model	AdjR^2	Model Rank	p-value	Model Rank	Variables Used
I	0.223	1	< 0.0001	3	CanopyCover CanopyCover^2 LargeTreeDensity ShrubDensity
II	0.221	2	< 0.0001	4	CanopyCover CanopyCover^2 LargeTreeDensity
III	0.203	3	< 0.0001	6	CanopyCover CanopyCover^2 ShrubDensity
IV	0.191	4	< 0.0001	5	CanopyCover GroundCover
V	0.191	5	< 0.0001	2	CanopyCover LargeTreeDensity
VI	0.189	6	< 0.0001	10	CanopyCover CanopyCover^2
VII	0.172	7	< 0.0001	8	CanopyCover
VIII	0.141	8	< 0.0001	1	LargeTreeDensity LargeTreeDensity^2
IX	0.133	9	< 0.0001	7	GroundCover
X	0.108	10	< 0.0001	9	LargeTreeDensity
XI	0.032	11	0.0055	11	ShrubDensity
XII	0.028	12	0.11	12	DeciduousBA

Table 17. Population Density Models: American Redstart

Predicted density (Territories/0.25 ha) is equal to D.

Model I	D = 0.122 - 0.00299(GroundCover) + 0.0000467(GroundCover)^2 - 0.00109(Ironwood/MapleCover)
Model II	D = 0.0598 - 0.00277(GroundCover) + 0.000051(GroundCover)^2
Model III	D = 0.023 + 0.00187(GroundCover) - 0.0012(Ironwood/MapleCover)
Model IV	D = -0.057 + 0.00262(GroundCover)
Model V	D = 0.149 - 0.00204(Ironwood/MapleCover)
Model VI	D = 0.238 - 0.00224(CanopyCover)
Model VII	D = 0.199 - 0.00326(LargeTreeDensity) + 0.0000126(LargeTreeDensity)^2
Model VIII	D = 0.136 - 0.00183(ShrubCover)
Model IX	D = 0.033 + 0.00000776(ShrubDensity)
Model X	D = 0.161 - 0.00159(SmallTreeDensity) + 0.00000572(SmallTreeDensity)^2
Model XI	D = 0.129 - 0.00073(LargeTreeDensity)
Model XII	D = 0.097 - 0.000185(SmallTreeDensity)

Tests of American Redstart Density Models

	Training Dataset		Verification Dataset		
Model	AdjR^2	Model Rank	p-value	Model Rank	Variables Used
I	0.143	1	0.0009	2	Ironwood/MapleCover GroundCover GroundCover^2
II	0.130	2	0.0003	1	GroundCover GroundCover^2
III	0.129	3	0.0039	5	Ironwood/MapleCover GroundCover
IV	0.100	4	0.0015	3	GroundCover
V	0.092	5	0.097	9	Ostrya/MapleCover
VI	0.050	6	0.039	8	CanopyCover
VII	0.035	7	0.0023	4	LargeTreeDensity LargeTreeDensity^2
VIII	0.024	8	0.098	10	ShrubCover
IX	0.016	9	0.023	7	ShrubDensity
X	0.015	10	0.31	12	SmallTreeDensity SmallTreeDensity^2
XI	0.013	11	0.006	6	LargeTreeDensity
XII	-0.006	12	0.18	11	SmallTreeDensity

Table 18. Population Density Models: Least Flycatcher

Predicted density (Territories/0.25 ha) is equal to D.

Model I D = -0.12 + 0.0122(DeciduousBA) + 0.00808(SubcanopyCover) - 0.0000797(Subcanopy)^2
Model II D = 0.006 + 0.0318(DeciduousBA) - 0.00126(DeciduousBA)^2
Model III D = 0.038 + 0.0146(DeciduousBA)
Model IV D = -0.146 + 0.011(DecidSubcanopy) - 0.000101(DecidSubcanopy)^2
Model V D = -0.029 + 0.0019(CanopyCover)

Tests of Least Flycatcher Density Models

	Training Dataset		Verification Dataset		
Model	AdjR^2	Model Rank	p-value	Model Rank	Variables Used
I	0.069	1	0.47 (-)	3	DeciduousBA SubcanopyCover SubcanopyCover^2
II	0.057	2	0.06 (-)	5	DeciduousBA DeciduousBA^2
III	0.051	3	0.10 (-)	4	DeciduousBA
IV	0.049	4	0.74 (-)	2	DecidSubcanopy DecidSubcanopy^2
V	0.015	5	0.87 (-)	1	CanopyCover

Results

1. Ovenbird. Ovenbird territories were spread densely and rather evenly throughout all plots. Although they are usually described as birds of the deciduous forest, they were actually slightly more common in pine-dominated stands on the plots. Although coniferous litter probably does not support as many invertebrate prey items as does deciduous litter (Smith & Shugart, 1987), the deciduous understory present in pine forests of Itasca Park may be sufficient for Ovenbirds to use those habitats without suffering food shortages. Possibly the preference for pine forests arises from their greater canopy height, which may trigger a habitat selection rule that the bird uses to select territories in mature forests. Ovenbirds were less likely to be present in open clearings and successional forests, which is consistent with the common description of them as forest-interior birds.

1.1. POP Models. Canopy Cover, Conifer Basal Area, Shrub Density, and Deciduous Basal Area were the variables used in the Ovenbird models. The rule-based model used only Canopy Cover, although the data would have justified the inclusion of Conifer Basal Area as a significant factor. I believed that, because Ovenbirds are typically associated with deciduous forest through most of their range, a model that assumed a positive relationship with conifers would probably not reflect the actual process by which Ovenbirds select habitat and would be likely to give misleading results. The regressions suggested that Ovenbirds were more common in closed-canopy forests, with a stronger preference for conifers than for deciduous trees. The curvilinear regression in Model E indicated a maximum abundance at about 85% Canopy Cover and 8000 shrubs/acre, and gave the best fit to the training data of the logistic models.

Both rule-based and logistic models performed poorly at distinguishing between used and unused sites. Only models including Conifer BA (Models A and E) performed even marginally better than the random null model. This apparent preference for coniferous forests was unlikely to reflect a generally applicable habitat preference because Ovenbirds are associated with deciduous forests over the majority

of their range. The failure of the models to detect differences in habitat quality may have been due to the extreme abundance of Ovenbirds in the park. Theoretical models of habitat selection (Fretwell and Lucas, 1969) suggest that as a population increases, marginal habitat is used more and more completely until all usable habitat is saturated. Because the Ovenbird population was high and territories were packed tightly throughout the plots, it is likely that much marginal habitat was being used and was not distinguishable from the optimal habitat by presence/absence measures. In particular, the successional aspen forest in the Sewage Lagoon plot was used by Ovenbirds although the vegetation was much younger and more open than that described as optimal habitat in the literature.

1.2. Population Density Models. In the regressions used to build these models, the Ovenbirds showed a fairly strong positive response to the presence of large pines (Models I through VI). Variables that were marginally significant in simple regressions included Large Tree Density, CanopyCover, and Ground Cover (Models VII, XI, XII). The highest R^2 value was achieved by a model combining White Pine BA, Red Pine BA, and Canopy Cover (Model I). The usefulness of a model relying on pine measurements is questionable on other grounds, however. Ovenbirds are described as preferring deciduous forests in many parts of their range, and the presence of pines is not likely to be the real factor to which Ovenbirds respond when selecting their territories. A model using Large Tree Density (Model VII) gave the highest R^2 values of those models that did not include pine measurements.

Testing these models against the verification dataset showed all of them to be significant in explaining variations in point density. This modelling technique appeared to be much more successful than the POP models. The most obvious area of suboptimal habitat, the successional Sewage Lagoon forest, was filled with Ovenbird territories that were distinctly larger than the territories in adjacent mature forest, leading to lower point density values. In this case, the density-based modelling technique was able to detect differences in density (and presumably in habitat quality), despite the effects of packing territories into marginal habitat that had kept the presence/absence measures used in POP models from being effective.

The tests of the models also revealed that any preference for pine forest by Ovenbirds was much less important than the R^2 values had initially suggested. The three models built exclusively on pine variables (IV to VI) were actually the least successful with independent data. The most successful model (IX) assumed only a preference for forests with large trees and a closed canopy, without including information about the tree species constituting the forest.

2. Red-eyed Vireo. Like the Ovenbirds, Red-eyed Vireos were abundant and rather evenly distributed throughout the plots. Almost all space on the plots was tightly packed with Vireo territories. The few areas with gaps in the distribution of Vireo territories were often characterized by the presence of clearings and relatively open woodland, but this was not a consistent pattern. Red-eyed Vireos were present in equal abundance in pine and deciduous forest.

2.1. POP Models. The rule-based model and two of the logistic regression models were built using Canopy Cover as the only variable. Deciduous BA was used in Logistic Model C. These were the only variables that were at all useful in separating used from unused sites in the training dataset. Model C gave a better fit to the data than did the other logistic regression models. Even so, all of the models performed poorly at distinguishing used from unused points in the verification dataset. This suggested the possibility that, as had happened with the Ovenbirds, the dense population was using suboptimal habitat and making it difficult for presence/absence measurements to characterize good habitat quality for the species.

2.2. Population Density Models. The major habitat response that was initially suggested by the development of these models from the training dataset was a preference for mature, closed forest. Red-eyed Vireos were more common in areas with high Canopy Cover and low Ground Cover (Models V, VI, VII), which indicated closed forest habitats. Large Tree Density was not significant in a linear simple regression (Model X), but showed a strong curvilinear response, suggesting that a moderately high number of large trees (c. 80/acre) was characteristic of the most favorable habitat (Models I and II). This appeared to be the major explanatory

variable, with Canopy Cover and Ground Cover being useful because of their correlation with it. Curvilinear transformations of other variables were not effective. A weak negative relationship with Small Tree Density also appeared, suggesting that an open subcanopy might be important to the Vireos. The signs of the variable coefficients in all models were consistent with a preference for mature, closed forest. Only structural criteria seemed to be important - distinctions between deciduous and coniferous trees or species-specific vegetation variables were not effective in explaining the distribution of Red-eyed Vireos in the training dataset. However, the R^2 values were low for all of these models, with a maximum value of 0.12 for Model I, and the model results should be viewed with caution.

The application of the models to the verification dataset underscored their weakness. Four of ten were negatively correlated with Red-eyed Vireo density on the verification points, and only one (Model X) was even marginally significant in predicting bird density. Model X indicated a positive linear relationship with Large Tree Density, and had had the worst R^2 values of all the Vireo models. The p-value of 0.07 may indicate a slight preference for mature forests with numerous large trees, but is not strong evidence, particularly since the models including Canopy Cover were very poor.

The failure of all of these models to successfully predict population density indicated that Red-eyed Vireos were broad habitat generalists. Any preference for particular habitat types was swamped by the effects of crowding vireo territories into all available space. Unlike the case of the Ovenbirds, the population density models for Red-eyed Vireos gave no particular improvement over presence/absence models in detecting preferred habitat. While the Ovenbirds had a definite preference for mature forest that was masked by their high abundance, there was no firm evidence that Red-eyed Vireos had any habitat preferences among the range of forest types represented in the study.

3. Pine Warbler.

Pine Warbler territories were clustered about the mature red pines and white

pines, and were virtually absent if there were no pines nearby. The birds could, however, be present where the pine trees were widely enough spaced to be missed by the 11 m vegetation sampling plots.

3.1. POP Models. Pine BA and Pine Proximity were the most important variables in the rule-based and logistic regression models. Both variables had a positive relationship with Pine Warbler abundance. Logistic Model B used the basal areas of red pines and white pines separately in curvilinear regressions, in an attempt to distinguish the effects of the two tree species on Pine Warbler distribution. Model A gave the closest fit to the training data of the logistic regression models.

The strong habitat specificity of this species allowed several of the models to distinguish used from unused sites in the verification dataset very clearly, with extremely low p-values. Logistic Model B performed very poorly, however, although it had had a fairly good fit to the training data. Apparently the complexities of fitting the curvilinear variables for the two pine species to a regression curve produced a model that reproduced the idiosyncracies of the training dataset well but did not reflect the biological realities of habitat use by Pine Warblers.

3.2. Population Density Models. The most effective variables in the multiple regressions from the training dataset were Pine Proximity and Red Pine BA, both of which were highly significant in single-variable regressions as well (Models IV and VI). White Pine BA was also significant in a simple regression (Model IX) but was only marginally significant when regressed together with Red Pine BA, and had a much lower regression coefficient (Model V). This suggested that the warblers might be distinguishing between the two pine species, and that white pines could constitute inferior habitat in some way. The highest R^2 value was reached with a model combining Red Pine BA and Pine Proximity (Model I). Curvilinear terms were not significant when added to the models.

When the models were applied to the verification dataset, the strong association of this species with pine trees was confirmed. However, the possibility of a species-specific association with red pine was not supported. The single-variable model for Red Pine BA (VI) was not a significant predictor of Pine Warbler density, although the

corresponding model for White Pine BA (IX) was - in other words the preference for pine species was reversed from the training dataset. There is no consistent evidence from this study for either species of pine being preferred over the other by Pine Warblers. The most effective models were those that did not distinguish between red pines and white pines but used only Pine Proximity and Pine BA as predictor variables. Model VIII, which used the Pine Cover variable, was not significant in the test against independent data. Pine Cover is probably not a very good index to the importance of pines in a forest, since pine foliage is borne at a great height on the tree and is often blocked from view by lower deciduous leaves when making cover measurements. Canopy Cover (Model X) was also non-significant, as was expected from the low R^2 value in the training dataset.

4. Common Yellowthroat. Yellowthroats were strongly associated with the small open marshes. A few areas in the cleared portion of the Sewage Lagoon plot that had few trees and a dense herbaceous ground cover also held yellowthroat territories. Territories were not found in areas of mature forest, however.

4.1. POP Models. The most important habitat variable for building Common Yellowthroat models from the training dataset was Marsh Proximity, which was used in three logistic regression models. Canopy Cover had a negative relationship with yellowthroat abundance and was also used in the logistic regressions. Model D, which used both Marsh Proximity and Canopy Cover, had the best fit to the training data of the logistic regression models. The rule-based model also used both of these variables. Rules were designed to allow an increased probability of finding yellowthroats in forests as a spill-over from nearby marshes, and to identify open areas in upland habitats as possible yellowthroat habitat.

The verification dataset suffered from an uneven representation of the major habitat types for this species, as all but one of the verification points were upland forest habitat and the single marsh point was the only one occupied by a yellowthroat territory. This was representative of the rarity of marshy sites on the census plots but did lead to extreme statistical results when the models were tested. A verification

dataset designed for a more effective test of Common Yellowthroat models would include more sites in marsh and clearing habitat. All of the models that included Marsh Proximity were highly effective in detecting high-quality habitat. The logistic regression model using only Canopy Cover was less successful because a number of unoccupied points had relatively open canopies.

There are several interesting points concerning these different Common Yellowthroat models. The most complex logistic model, Model A, was the least effective at predicting abundance. It includes both linear and curvilinear terms for Marsh Proximity and Canopy Cover, but ends up predicting inaccurately low probabilities for marshy sites. The approach of simply including all of the relevant variables and curvilinear terms in a single regression equation does not necessarily improve predictive power and may actually make it worse.

Model D, which used Marsh Proximity and Canopy Cover, predicted high probabilities for marshy habitat and very low probabilities for upland forests. This model conformed quite closely to the verification data. However, the success of Model D is quite "brittle" and I believe it is unreliable. This model's predicted probabilities for forested sites are fantastically low (in the 10^{-15} range). While upland forests are certainly not good yellowthroat habitat, the occasional presence of small clearings or dense ground cover is probably still enough to allow a 1-2% chance of finding a yellowthroat on a forest site. It happened that there were none present on the verification sites during 1991, but if any had been present (not an unlikely event) the model would have been swiftly invalidated. The model should be flexible enough to accommodate the occasional presence of a bird in suboptimal habitat.

Model B, which used only Marsh Proximity, was the best of the logistic models, taking into account the most important habitat factor and remaining flexible enough to allow for a few birds being present on the uplands. However, its failure to include the effects of Canopy Cover is inconsistent with the observed data. There is definitely a difference between forested uplands and open, shrubby uplands as yellowthroat habitat. The logistic regression technique doesn't seem to give a good middle ground between Model B and Model D, which would allow a role for the Canopy Cover variable

without making it impossible for yellowthroats to be found in the forest. The rule-based Common Yellowthroat model seems to be the best, conforming closely to the verification dataset without leading to other unlikely results.

4.2. Population Density Models. The process of building these models from the training dataset showed Marsh Proximity to be the single most important habitat variable for yellowthroats, with a strong positive effect on their abundance (Model VII). The most useful way of incorporating the effects of the canopy layer was a curvilinear transformation of Tree Density, which indicated a steep decline in yellowthroat density with increasing numbers of trees, reaching a minimum at approximately 225 trees/acre (Model VIII). Canopy Cover (Model IX) indicates a similarly negative effect from the presence of trees, but does not give as high an R^2 value. Combining the Marsh Proximity and Tree Density variables gives the most effective model (Model I). Ground Cover was significant in a simple regression (Model XI), but was not helpful when combined with other variables.

When applied to the verification dataset, all of the models that included Marsh Proximity were highly significant as predictors of yellowthroat density, while the other models were not significant. This was consistent with what had been expected from the training dataset. The choice of variables for fine-tuning of the models was somewhat different from what had been expected, however. The more complex models that included variations of Tree Density (I, II, III, IV) and had had high R^2 values were not as effective as Model V, which included only Marsh Proximity and Canopy Cover.

5. Black-throated Green Warbler. These warblers appeared to be most common in mature pine forest, but were not strictly limited to them, occurring in deciduous forest as well. There were some discrepancies between plots in the occurrence of this species; it was completely absent from the mature forest of the Sewage Lagoon plot although there appeared to be suitable habitat present.

5.1. POP Models. The rule-based model indicated the most favorable habitat for Black-throated Green Warblers to be closed forest with some pine trees present. The logistic regression models gave a similar picture, with slight positive relationships to

Pine Proximity, Pine BA, and Canopy Cover. Model A, using Pine Proximity and Canopy Cover, gave the closest fit to the training data of the logistic regression models.

None of the models was significantly better than random at distinguishing used from unused habitat in the verification dataset - the rule-based model was most effective at $p = 0.17$. The categorical p-values of c. 0.2 were suggestive of a preference for closed forests with coniferous trees nearby, which is consistent with the habitat use described in the literature. Black-throated Green Warblers were not common enough to be packing the available habitat with their territories as did Red-eyed Vireos and Ovenbirds, and there were apparently suitable areas of pine forest that were completely unoccupied, so the failure to detect significant habitat responses does not seem to be caused by high populations masking distinctions in habitat quality. The preference for pine-dominated sites simply appears to be weak. Possibly there was some extraneous factor that caused differences between plots - as noted above, the Sewage Lagoon plot contained no Black-throated Green Warblers at all although it seemed to be good habitat.

5.2. Population Density Models. Single-variable regressions on the training dataset showed the Pine Proximity and Pine BA variables were significant explanatory variables, and Canopy Cover was marginally significant (Models VI, IX, X). Red Pine BA appeared to have a much stronger effect than White Pine BA (Model VIII). Balsam Fir BA (Models I and III) was actually associated more strongly with Black-throated Green Warblers than any of the other variables, giving the closest fit to the training dataset in Model I, but this was probably not biologically meaningful. Firs were small, uncommon understory trees on the plots, and were unlikely to have a strong influence on a canopy-dwelling species like this warbler. The association was probably coincidental - even the best of the Black-throated Green Warbler models had very low R^2 values. The most effective multiple regression model that was likely to have biological meaning combined Pine Proximity, Red Pine BA, and Canopy Cover (Model II).

Unfortunately, almost all of the models performed very poorly at predicting Black-throated Green Warbler density in the verification dataset. The preference for

pine-dominated forests that was suggested by the training dataset was very weak in the verification data. Four of the models gave predictions that were negatively correlated with Black-throated Green Warbler density. Skepticism about the value of Balsam Fir BA as a predictive variable turned out to be justified by the poor showings of those models that included it. Model X, which had the greatest predictive power and was the only model with a significant positive relationship to the verification dataset, used only a single structural vegetation variable (Canopy Cover). This model indicated a preference for closed-canopy, mature forest.

6. Chestnut-sided Warbler. These warblers were abundant in the successional forest on the southern half of the Sewage Lagoon plot. They were locally present in forest openings through the other plots.

6.1. POP Models. The rule-based model designated favorable habitat for Chestnut-sided Warblers to be relatively open forest without mature deciduous trees. This reflected their abundance in clearings and young forest, and allowed for the fact that some open pine forests also had moderate numbers of the species. Variables included in the logistic regression models included Canopy Cover, Deciduous BA, and Shrub Density. The models suggested a preference for low Canopy Cover and Deciduous BA, and high Shrub Density. The curvilinear regressions in Models B and E indicated that the most favorable habitat had about 30% Canopy Cover, which seems to be correct in excluding the treeless marshes from the category of optimum Chestnut-sided Warbler habitat. Model E, which included Canopy Cover and Shrub Density, had the best fit to the data of the logistic regression models.

Nearly all of the habitat models performed significantly better than random when applied to the verification dataset. The rule-based model was the most effective at distinguishing used from unused points, with a very low p-value. The least effective was Logistic Model C, which used only Deciduous BA as a variable. The dominance of deciduous trees, considered in the absence of other variables, apparently does not affect Chestnut-sided Warbler abundance. On the other hand, the strong habitat specificity of these warblers allowed the models that took into account the openness of

the forest (Canopy Cover) to identify favorable habitat quite successfully.

6.2. Population Density Models. A curvilinear regression of Canopy Cover was significant and showed about 20% cover to be most favorable for Chestnut-sided Warblers (Model VI). Areas with low Deciduous BA and low Large Tree Density were favored for similar reasons, both factors being associated with open, early successional forests (Models X and XII). The curvilinear regression of Large Tree Density indicated that the warbler density dropped off rapidly with increasing numbers of large trees, reaching a minimum at c. 110 trees/acre. Shrub Density was positively associated with dense Chestnut-sided Warbler populations (Model XI). The model with the highest adjusted R^2 used Shrub Density, Large Tree Density, and the curvilinear Canopy Cover terms (Model I).

Nearly all of the models were highly effective at predicting bird density in the verification dataset and had extremely small p-values. As in the POP models, Deciduous BA (Model XII) was the only variable used that was not significant as a predictor. While Canopy Cover had been the most important variable in the training dataset, the most effective models at predicting the verification dataset were the ones that incorporated Large Tree Density as a variable. This was a minor disagreement, however, since both variables are similar measures of forest maturity and the models were all highly significant.

7. American Redstart. Redstarts were found primarily in the young aspen forest on the Sewage Lagoon plot, but also were occasionally present in relatively open areas of mature forest. Their habitat preferences seem similar to those of Chestnut-sided Warblers, but the requirement of redstarts for open forest is not so crucial.

The observation of Sherry (1979) that American Redstarts tend to be excluded from areas occupied by Least Flycatcher colonies was not obviously confirmed in this study. On Sherry's sites in New Hampshire, redstarts were abundant enough that "holes" in their distribution around flycatcher colonies were obvious. In Itasca Park neither species occupies a large enough proportion of the area for any interaction to be clear, however, and there are occasional overlaps between territories of the two species.

7.1. POP Models. The rule-based model for redstarts indicated their preferred habitat to be areas with dense ground cover and little canopy cover of ironwood and sugar maple overhead. Small ironwood and sugar maple trees formed a dense, spreading subcanopy in the mature forests where Redstarts were unlikely to be found, and the combined cover of these two tree species seemed to be a better criterion than simple Canopy Cover. The amount of ground cover seems to be unlikely as a factor in habitat choice by Redstarts, which are not ground-dwelling birds, but there was a strong correlation between Ground Cover and Redstart presence in the data. Open forests do tend to have a more dense herb layer, but the Ground Cover/Redstart density correlation was not fully explained by variations in Canopy Cover. Ground Cover and Ironwood/Maple Cover were selected as the explanatory variables for the rule-based model.

The logistic regression models included Ground Cover, Canopy Cover, Shrub Density, and Ironwood/Maple Cover as variables, with Models D and E giving the best fit to the training dataset. When evaluated against the verification dataset, the models generally had marginal significance on the categorical test, with p-values ranging from 0.08 to 0.12. The rule-based model had the best classification accuracy. Only Model D had a high enough p-value to fail unambiguously. The results suggested that Ground Cover, Canopy Cover, and Ironwood/Maple Cover were probably useful predictors, and that Shrub Density was unreliable.

7.2. Population Density Models. Multiple regression on the training dataset suggested that Ground Cover and Ironwood/Maple Cover were the variables that best explained Redstart density (Models I to V). Other variables used, in models that had lower adjusted R^2 values, included Shrub Density, Large Tree Density, Small Tree Density, and Canopy Cover. Curvilinear regressions suggested that the least favorable habitat was characterized by a Ground Cover of c. 30% (Model II) and Large Tree Density of c. 130 trees/acre (Model VII). The model with the highest adjusted R^2 value was Model I, which included Ironwood/Maple Cover and curvilinear Ground Cover terms.

The application of these models to the verification dataset generally confirmed the habitat relationships noted above. The models that had been expected to be most effective (I and II) were confirmed at very low p-values.

8. Least Flycatcher. The semi-colonial distribution pattern of this species, which was noted by Sherry and Holmes (1985), was also apparent on the census plots in this study. Clumps of flycatcher territories were present at the Peace Pipe, Checkpoint, and Sewage Lagoon plots. Least Flycatchers are described in the literature as preferring open deciduous or mixed forests. However, it was quite difficult to identify any habitat characteristics on the plots that were clearly effective in separating used from unused habitat, as the vegetation in the flycatcher colonies varied from extremely open deciduous forest to mature pine forest with a dense understory of maple.

8.1. POP Models. The training dataset suggested a slight preference for closed-canopy forests with large deciduous trees. The difference between using Canopy Cover and Subcanopy Cover in the analysis was slight, but the fact that this species typically feeds within 10 meters of the ground suggested that the nature of the subcanopy might be the determining factor. The rule-based model was therefore built to reflect a preference for a closed deciduous subcanopy. Logistic regression models included Canopy Cover and Deciduous BA as variables. Model B, which used the curvilinear Canopy Cover terms, gave the best fit to the training dataset.

Both the rule-based and logistic regression models performed very poorly at distinguishing used from unused sites in the verification dataset. Categorical p-values were in the range of 0.88-0.99, which is considerably worse than a random model.

8.2. Population Density Models. None of these multiple regression models explains very much of the variation in flycatcher density, with the highest adjusted R^2 being 0.07. Only five models were built, because very few habitat variables could be found with any significant relation to flycatcher density in the training dataset. Deciduous BA did have a significant positive effect in the training dataset (Model III),

which is consistent with the findings of Sherry and Holmes (1985) that Least Flycatchers preferred deciduous forest. A curvilinear regression of Deciduous Subcanopy Cover (Model IV) was also significant, indicating that intermediate levels of subcanopy cover (c. 50%) were most favorable. Canopy Cover, Shrub Density, and Tree Density had no significant relationship to bird density. The model with the highest R^2 (Model I) included Deciduous BA and the curvilinear transformation of Subcanopy Cover.

Even these rather shaky relationships fell apart when the models' predictions were compared to observed flycatcher density in the verification dataset. None of the models even had a positive relationship with flycatcher density, and Model II came close to having a significant negative relationship. The social dynamics of Least Flycatcher colonies appear to have a much stronger influence on their distribution in the park than do habitat preferences. Any description of the preferred habitat for this species in Itasca Park would be inconclusive.

Discussion

1. Do the Models Indicate Habitat Quality?

Strictly speaking, this question cannot be answered with these data. I made no direct evaluations of habitat quality or correlates such as nesting success, food availability, predation rates, or mating status of territorial males — only the data on habitat structure, presence/absence, and point density were used. However, the density models that were most significant in the test against independent data were in every case consistent with the characterization from the literature of typical habitat for that species, and I have little reason to doubt that population density is related to habitat quality. (Probability of presence appears to be less strongly related to habitat quality: it was less successful than density measures in distinguishing used from unused sites for abundant species like the Ovenbird.) Only territorial birds were censused, preventing subdominant "floaters" in low-quality habitat from confusing the relationships. Since multiple visits were made to the plots, differences in detectability between mated and unmated territorial males were unlikely to strongly affect the territory maps. Territories are almost always rarer and/or larger in habitats that would be expected to be of lower quality (e.g., open forests for Ovenbirds; forests with few or no pines for Pine Warblers). The only case in which high population levels might be masking habitat quality trends is the Red-eyed Vireo. Vireos show no clear habitat preference from the models. However, they are abundant enough that less competitive territorial males might be forced into unfavorable habitats and end up being packed in at a population density equalling or exceeding the density in high-quality habitat. It is not possible to tell whether this is in fact the case, or whether the Vireos really have no preferences among the habitat types in the study. In general, however, density models that are effective in predicting verification data appear to be identifying high-quality habitat for the appropriate species, and would be useful in making land management decisions for conservation of these species.

2. Statistical problems in Regression Analysis.

Regression models include assumptions of homogeneous variance and independence of habitat variables that are violated at several points in the data analysis.

The assumption of homogeneous variance leads to a particularly intractable problem in the regression models of population density. If, for example, Chestnut-sided Warbler point density is graphed against Canopy Cover, there is a wide scatter in density values at low to moderate Canopy Cover, but at higher Canopy Cover the density values are almost uniformly zero. Similar problems occur for several other species, including Common Yellowthroat, Least Flycatcher, and Pine Warbler. Habitat that is definitely not suitable is rarely occupied, while suitable habitat has a wide range of density values. (Not all of the suitable habitat will be occupied, and the chance location of sampling points on the edge or center of a territory will give rise to random variations in point density values and lower R^2 values. Measuring point density values over a greater area than 0.25 ha would even out this variation but would make the particular vegetation measurements taken at the sampling point less representative of the larger area.) The very fact that a variable is useful in explaining the bird's distribution leads to non-constant variance. Several variable transformations, on both the X and Y axes, were tested, but they were uniformly unsuccessful in alleviating the problem. Models that were useful in predicting which sites would be heavily used were generally retained in spite of heterogeneous variance. This sometimes led to certain sites, presumed to have low habitat quality, giving a negative predicted density as the regression line passed below the x axis. While negative densities are obviously not possible in reality, the models were still useful if viewed as a relative index of habitat quality, and their predictions were often strongly related to the independent data.

A second persistent problem in the regression models was the inclusion of intercorrelated habitat variables. Many of the important habitat variables were related to one another. These included pairs of variables such as: Marsh Proximity-Canopy Cover; Pine Proximity-Pine BA; Pine BA-Large Tree Density; and Canopy Cover-Ground Cover. These variables conveyed some unique information about the habitat but were also correlated with other variables. Many highly correlated variable pairs

were excluded from the analysis. However, in some cases the unique information provided by both variables seemed potentially useful because of a markedly better fit to the training data and/or the biological importance of the variables, and the model was retained for testing against independent data (e.g., both Marsh Proximity and Canopy Cover were important to Yellowthroats; both Pine Proximity and Pine BA were important to Pine Warblers). The inclusion of intercorrelated variables is a less serious problem in situations where: a) the model is not applied beyond the range of habitat types used in building the model; and b) the objective is to predict the value of the response variable rather than to determine which predictors have a significant causative influence on the response. This analysis fits both descriptions. Sometimes models with intercorrelated variables were highly successful in predicting the verification dataset. In these cases, I believe that the practical usefulness of the models in distinguishing habitat quality justifies the somewhat questionable statistical methods used to derive them. The practice of using independent verification data to validate the models is extremely useful for cutting through debate about relationships between the model and the training dataset and focusing attention on the actual predictive power of the model.

3. Choosing among Regression Models.

A major criterion used in initial evaluation of the population density models built by multiple regression was the value of the adjusted R^2 statistic, which indicated how well the model fit the training dataset. (The simple R^2 statistic was not as useful since its value was dependent on the number of variables used in the model.) It is instructive to compare the rankings of models by the adjusted R^2 for the training dataset with the rankings by p-value for the verification dataset. Often the models that had the closest fit to the training dataset were not the most effective at predicting the training dataset. It is possible for models to suffer from overfitting, in which the correlation between random variation in predictor variables and the dependent variable is scavenged to improve the fit with the training data. If much of the correlation in the model stems from such random effects and little of it is biologically meaningful, the model will perform poorly against independent data although it may do well at "predicting" the

training data. It is difficult to know whether a model is overfitted before testing it against verification data, but many of the overfitted models seem to be complex, incorporating several variables and curvilinear terms that are not related in ways that have obvious biological significance. It is well to be cautious about complicated models that go beyond the variables that are known to be important.

This should underline the importance of having independent data available with which to evaluate any model. If no verification data had been used, and the only way of evaluating the models had been the R^2 statistic, many of the models chosen as being most effective would not actually have been the best ones. Testing models against independent data confirms that they do in fact have wider applicability beyond the dataset used to build them.

4. Population Density Models vs. POP Models.

Several versions of these two types of models were developed for each of the eight species and tested against the verification dataset. The best indicators of the accuracy with which they distinguished used from unused habitat were the p-value for the categorical test, derived from the analysis program (for POP models) and the variable coefficient p-value (for population density models). Because the hypotheses being tested with these p-values are somewhat different, the tests are not strictly comparable, but for several species it seems clear that the population density models are more effective in identifying preferred habitat. The POP models never identified significant habitat preferences for the Ovenbird and American Redstart, but density models showed strong relationships that were consistent with results from the literature. The Black-throated Green Warbler and Chestnut-sided Warbler also seem to show habitat preferences more strongly in density models than in POP models. On the other hand, the habitat preferences of Yellowthroats and Pine Warblers were effectively detected by both model types, and Red-eyed Vireos and Least Flycatchers failed to show significant preferences for either type. The problem with the Least Flycatcher was probably related to its colonial territory system, which apparently had a stronger influence on distribution than did habitat preferences. While the POP models suggested

that both of the most abundant species (Red-eyed Vireo and Ovenbird) were habitat generalists, the density models detected the differences in habitat use between the two species, showing Ovenbirds to prefer mature forest and confirming only the Vireo as a generalist species that can effectively exploit diverse habitat types.

5. Roles of Data and Experience in POP Models.

The three methods used to develop POP models make differing assumptions about the relationship between habitat type and the probability of the site being occupied. The logistic regression models give a linear or curved continuous relationship between the habitat variables and probability of presence. The rule-based models, on the other hand, make the relationship a step-like function, with abrupt discontinuities between regions of uniform abundance. Fuzzy models are similar to the simple rule-based models, but give a gradual transition between different abundance levels rather than an abrupt one. I am not certain which of the methods is a better reflection of the way that a bird population arranges itself across habitat types. Perhaps, as rule-based or fuzzy models would imply, there are psychological rules-of-thumb, either sharp or blurred, that the birds use to evaluate the suitability of a habitat. A continuous linear function such as is produced by logistic regression may still be appropriate, however. There is no a priori reason to believe that one method will be more successful at prediction than another.

In general, the classification accuracies of rule-based, fuzzy, and logistic regression modelling methods do not differ dramatically. No one modelling method consistently gives the best p-value when tested against verification data. This even showing may actually be slightly favorable to the rule-based and fuzzy methods, since three to five logistic regression models were built for each species, while only one rule-based model and one fuzzy model were built. There are no clear differences, however. Fuzzy models did not consistently give results that improved over the results of rule-based models. The choice of which modelling method to use would have to be made on non-statistical grounds, therefore.

Using logistic regression techniques by computer allows models to be rapidly

and easily built, and the importance of many variables can be examined in a short period of time. If one is relying primarily on the data for information about habitat use, without relevant personal experience to draw upon, running a few logistic regression models will certainly give some useful "best guesses". Statistical indices such as coefficient p-values and deviance values give a numerical basis for deciding which models are more useful in explaining the data. It can be risky to use goodness-of-fit statistics like deviance as the only criterion for selecting a model, on the other hand. If there is no biological relevance to the variables included, the model may be less applicable in situations outside the training dataset.

Building rule-based or fuzzy models, on the other hand, can be rather slow and tedious. Graphs of variables and presence/absence data are examined individually for potentially important relationships. Because of the time-consuming nature of the process, however, it does give a more intimate understanding of the data than does a swift logistic regression. Rules are decided upon rather than automatically drawing a best-fit line. The rule may not exactly reflect what the data say, which can be a source of error. On the other hand, the data will inevitably incorporate some degree of sampling error and give a somewhat erroneous estimate of probability of presence, so the rule-based models' partial independence from the data may allow the effects of sampling error to be corrected. The process of specifying the rules allows personal judgements about the habitat relationships to be incorporated into the model. This is particularly useful in cases where the data are rather sketchy, as the model-builder can draw on personal experience with the species to round out the picture. (The Common Yellowthroat models are a good example, where the data available were limited because the species was uncommon and highly selective in habitat use. The regression models gave extremely unlikely results, and the rule-based model was the only one that appeared trustworthy.) Of course, this does throw back more of the responsibility on the modeller as expert. Presumably the modeller is knowledgeable about the species, but if this is not the case, then personal experience is of little help.

If the model is incorporated into an expert system, the rule-based method has a great advantage in that its "If-Then" logic can easily be translated into expert system

format. The reasoning used to derive rules can be incorporated into the system in lay terms. Logistic regression models are hard to transfer to this format, and the rationale for model selection is necessarily statistical. If the use of a model by non-experts is important, the relative ease of understanding a rule-based model may give it more credibility.

In general, regression techniques are favorable for situations where rapid examination of many variables is important, and where it is unlikely that personal judgement will be able to improve on the information present in the data. Rule-based and fuzzy models are favorable when accurate expert knowledge is available, when the data are sparse or unreliable, or when it is important that the rationale used to build the model be accessible to the user, as in an expert system.

Literature Cited

Anderson, S.H., and H.H. Shugart. 1974. Habitat selection of breeding birds in an east Tennessee deciduous forest. Ecology 55:828-837.

Askins, R.A., J.F. Lynch, and R. Greenberg. 1990. Population declines in migratory birds in eastern North America. Current Ornithology 7:1-58.

Askins, R.A. and M.J. Philbrick. 1987. Effect of changes in regional forest abundance on the decline and recovery of a forest bird community. Wilson Bull. 99: 7-21.

Bart, J., D.R. Petit, and G. Linscombe. 1984. Field evaluation of two models developed following the Habitat Evaluation Procedures. Trans. N.A. Wildl. Nat. Res. Conf. 49: 489-499.

Blenden, M.D., M.J. Armbruster, T.S. Baskett, and A.H. Farmer. 1984. Evaluation of model assumptions: The relationship between plant biomass and arthropod abundance. Pp. 11-15 in: "Wildlife 2000" (J. Verner, M.L. Morrison, and C.J. Ralph, eds.). Univ. Wisconsin Press, Madison, Wisconsin.

Breckenridge, W.J. 1956. Measurements of the habitat niche of the Least Flycatcher. Wilson Bull. 68: 47-51.

Brennan, L.A., W.M. Block, and R.J. Gutierrez. 1986. The use of multivariate statistics for developing Habitat Suitability Index models. Pp. 177-182 in: "Wildlife 2000" (J. Verner, M.L. Morrison, and C.J. Ralph, eds.). Univ. Wisconsin Press, Madison, Wisconsin..

Brittingham, M.C., and S.A. Temple. 1983. Have cowbirds caused forest songbirds to decline? BioScience 33:31-35.

Brush, T. and E.W. Stiles. 1986. Using food abundance to predict habitat use by birds. Pp. 57-64 in: "Wildlife 2000" (J. Verner, M.L. Morrison, and C.J. Ralph, eds.). Univ. Wisconsin Press, Madison, Wisconsin.

Capen, D.E., J.W. Fenwick, D.B. Inkley, and A.C. Boynton. 1986. Multivariate Models of Songbird Habitat in New England Forests. pp. 171-177 in: "Wildlife 2000" (J. Verner, M.L. Morrison, and C.J. Ralph, eds.). Univ. Wisconsin Press, Madison, Wisconsin.

Carrick, R. 1963. Ecological significance of territory in the Australian Magpie. Proc. 13th Intern. Ornithol. Congr. 740-753.

Cole, C.A., and R.L. Smith. 1983. Habitat Suitability Indexes for monitoring wildlife populations: an evaluation. Trans. N.A. Wildl. Nat. Res. Conf. 48: 367-375.

Collins, S.L. 1983a. Geographic variation in habitat structure for the wood warblers in Maine and Minnesota. Oecologia 59:246-252.

Collins, S.L. 1983b. Geographic variation in habitat structure of the Black-throated Green Warbler. Auk 100: 382-389.

Collins, S.L., F.C. James, and P.G. Risser. 1982. Habitat relationships of wood warblers (Parulidae) in northern central Minnesota. Oikos 39:50-58.

Connell, J.H. 1983. On the prevalence and relative importance of interspecific competition: Evidence from field experiments. Amer. Nat. 122:661-696.

Cornell Laboratory of Ornithology. 1989. An introduction to the Breeding Bird Census. 7 page mimeo.

Coulson, R.N., L.J. Folse, and D.K. Loh. 1987. Artificial intelligence and natural resource management. Science 237: 262-267.

Diamond, J.M. 1975. Assembly of species communities. Pp. 342-444 in: Ecology and Evolution of Communities (M.L. Cody and J.M. Diamond, eds.). Belknap, Cambridge, Massachusetts, USA.

Ehrlich, P.R., D.S. Dobkin, and D. Wheye. 1988. The Birder's Handbook. Simon & Schuster, New York, New York, USA.

Ficken, R., M. Ficken, and D.H. Morse. 1968. Competition and character displacement in sympatric pine-dwelling warblers. Evolution 22:307-314.

Finch, D.M. 1989. Habitat use and habitat overlap of riparian birds in three elevational zones. Ecology 70:866-880.

Fretwell, S.D., and H.L. Lucas. 1969. On territorial behavior and other factors influencing habitat distribution in birds: I. Theoretical Development. Acta Biotheoret. 19: 16-36.

Giarratano, J., and G. Riley. 1989. Expert Systems: Principles and Programming. PWS-Kent Publishing Co., Boston, Mass.

Gibbs, J.P., and J. Faaborg. 1990. Estimating the viability of Ovenbird and Kentucky Warbler populations in forest fragments. Cons. Biol. 4: 193-196.

Greenberg, R. 1979. Body size, breeding habitat, and winter exploitation systems in Dendroica. Auk 96:756-766.

Hinde, R.A. 1956. The biological significance of the territories of birds. Ibis 98: 340-369.

Huxley, J. 1934. A natural experiment on the territorial instinct. Brit. Birds 27: 270-277.

James, F.C. 1971. Ordination of habitat relationships among breeding birds. Wilson Bull. 83:215-236.

James, F.C. 1978. On understanding quantitative surveys of vegetation. American Birds 32: 18-21.

James, F.C., and C.E. McCulloch. 1990. Multivariate analysis in ecology and systematics: Panacea or Pandora's box? Annu. Rev. Ecol. Syst. 21: 129-166.

James, F.C., and H.H. Shugart Jr. 1970. A quantitative method of habitat description. Audubon Field Notes 24: 727-736.

James, F.C., and N.O. Wamer. 1982. Relationships between temperate forest bird communities and vegetation structure. Ecology 63:159-171.

Kluyver, H.N. and L. Tinbergen. 1953. Territory and the regulation of density in titmice. Arch. Neerl. Zool. 10: 265-287.

Lancia, R.A., S.D. Miller, and D.A. Adams. 1982. Validating habitat quality assessment: an example. Trans. N.A. Wildl. Nat. Res. Conf. 47: 96-110.

Lancia, R.A., D.A. Adams, and E.M. Lunk. 1986. Temporal and spatial aspects of species-habitat models. pp. 65-70 in: "Wildlife 2000" (J. Verner, M.L. Morrison, and C.J. Ralph, eds.). Univ. Wisconsin Press, Madison, Wisconsin.

Latka, D.C. and J.W. Yahnke. 1986. Simulating the roosting habitat of Sandhill Cranes and validating suitability-of-use indices. pp. 19-22 in: "Wildlife 2000" (J. Verner, M.L. Morrison, and C.J. Ralph, eds.). Univ. Wisconsin Press, Madison, Wisconsin.

Laymon, S.A. and R.H. Barrett. 1986. Developing and testing Habitat-Capability models: Pitfalls and recommendations. pp. 87-92 in: "Wildlife 2000" (J. Verner, M.L.

Morrison, and C.J. Ralph, eds.). Univ. Wisconsin Press, Madison, Wisconsin.

MacArthur, R.H., and J. MacArthur. 1961. On bird species diversity. Ecology 42:594-598.

Maurer, B.A. and R.C. Whitmore. 1981. Foraging of five bird species in two forests with different vegetation structure. Wilson Bull. 93:478-490.

Morrison, D.G. 1969. On the interpretation of discriminant analysis. J. Marketing Research 6: 156-163.

Morse, D.M. 1985. Habitat Selection in North American Parulid Warblers. In: "Habitat Selection in Birds" (M. L. Cody, Ed.). Academic Press, inc.

Mountainspring, S., and J.M. Scott. 1985. Interspecific competition among Hawaiian forest birds. Ecological Monographs 55: 219-239

Noon, B.R., D.K. Dawson, D.B. Inkley, C.S. Robbins, and S.H. Anderson. 1980. Consistency of habitat preference of forest bird species. Trans. N.A. Wildl. Nat. Res. Conf. 45: 226-244.

O'Connor, R.J. 1981. Habitat correlates of bird distribution in British census plots. Studies in Avian Biology 6:533-537

Orians, G.H., and M.F. Willson. 1964. Interspecific territories of birds. Ecology 45:736-745.

Parmelee, D.F. 1977. Annotated checklist of the birds of Itasca State Park and surrounding areas. Loon 49: 81-95.

Peterson, R.T. 1980. A field guide to the birds. 3rd Edition.

Robbins, C.S. 1970. An international standard for a mapping method in bird census work recommended by the International Bird Census Committee. Audubon Field Notes 24:722-726.

Robinson, S.K. and R.T. Holmes. 1984. Effects of plant species and foliage structure on the foraging behavior of forest birds. Auk 101:672-684.

Rotenberry, J.T. 1985. The role of habitat in avian community composition: physiognomy or floristics? Oecologia 67: 213-217.

Schneider, M. and A. Kandel. 1991. General purpose fuzzy expert systems. Ch. 2 in: Fuzzy Expert Systems [A. Kandel, Ed..]. CRC Press, Boca Raton, Florida.

Schroeder, R.L. 1982. Habitat Suitability Index models: Pine Warbler. Western Energy and Land Use Team, OBS, U.S. Fish & Wildlife Service.

Sherry, T.W. 1979. Competitive interactions and adaptive strategies of American Redstarts and Least Flycatchers in a northern hardwoods forest. Auk 96:265-283.

Sherry, T.W., and R.T. Holmes. 1985. Dispersion patterns and habitat responses of birds in northern hardwoods forests. In "Habitat Selection in Birds" (M.L. Cody, ed.). Academic Press, Inc.

Smith, T.M. and H.H. Shugart. 1987. Territory size variation in the Ovenbird: the role of habitat structure. Ecology 68: 695-704.

Starfield, A.M. 1990. Qualitative, rule-based modelling. BioScience 40: 601-604.

Starfield, A.M., B.P. Farm, and R.H. Taylor. 1989. A rule-based ecological model for the management of an estuarine lake. Ecol. Model. 46:107-119.

Stenger, J. and J.B. Falls. 1959. The utilized territory of the Ovenbird. Wilson Bull. 71: 125-140.

Stewart, R.E. and J.W. Aldrich. 1951. Removal and repopulation of breeding birds in a spruce-fir community. Auk 68: 471-482.

Terborgh, J. 1990. Where have all the birds gone? Princeton Univ. Press, Princeton, NJ.

Temple, S.A. 1984. Predicting impacts of habitat fragmentation on forest birds: A comparison of two models. Pp. 301-304 in: "Wildlife 2000" (J. Verner, M.L. Morrison, and C.J. Ralph, eds.). Univ. Wisconsin Press, Madison, Wisconsin.

Tompa, F.S. 1962. Territorial behavior: the main controlling factor of a local Song Sparrow population. Auk 79: 687-697.

Van Horne, B. 1983. Density as a misleading indicator of habitat quality. J. Wildl. Manage. 47: 893-901.

Verner, J. and K.A. Milne. 1990. Analyst and observer variability in density estimates from spot-mapping. Condor 92: 313-325.

Wiens, J.A. 1977. On competition and variable environments. American Scientist 65: 590-597.

Wiens, J.A. 1981. Scale problems in avian censusing. Studies in Avian Biology 6: 513-521.

Wiens, J.A., and J.T. Rotenberry. 1981. Habitat Associations and community structure of birds in shrubsteppe environments. Ecol. Monographs 51:21-41.

Wilcove, D.S. 1985. Nest predation in forest tracts and the decline of migratory songbirds. Ecology 66: 1211-1214.

Zadeh, L.A. 1965. Fuzzy sets. Inf. Control 8:338-350.